

**VŠB – Technická univerzita Ostrava
Fakulta elektrotechniky a informatiky
Katedra informatiky**

**Metodika a nástroj na ověření procesů
Methods and Tools for Process Validation**

2017

Bc. Miroslav Stopka

VŠB - Technická univerzita Ostrava
Fakulta elektrotechniky a informatiky
Katedra informatiky

Zadání diplomové práce

Student: **Bc. Miroslav Stopka**
Studijní program: N2647 Informační a komunikační technologie
Studijní obor: 2612T025 Informatika a výpočetní technika
Téma: Metodika a nástroj na ověření procesů
Methods and Tool for Process Validation

Jazyk vypracování: čeština

Zásady pro vypracování:

Cílem diplomové je zpracování process mining části metodiky a její implementace ve formě pluginu do nástroje ProM nebo samostatně s napojením na ProM. Metodika se týká modelování, simulace a zpětné kontroly prováděných procesů. Nástroj bude, mimo jiné, podporovat zvolený metamodel, a dle něj identifikovat a vyhodnocovat reálné průběhy procesů.

Práce bude obsahovat zejména:

1. Seznámení s problematikou a metodikou, dopracování dalších částí týkající se process miningu.
2. Podrobný popis process mining části a její začlenění do nástroje.
3. Implementace nástroje
4. Experimenty, vyhodnocení

Seznam doporučené odborné literatury:

- [1] John F. Sowa, Knowledge Representation: Logical, Philosophical, and Computational Foundations, Brooks Cole Publishing Co., Pacific Grove, CA, ©2000
- [2] Alec Sharp, Patrick McDermott: Workflow Modeling: Tools for Process Improvement and Application Development, Artech House; 2 edition (October 31, 2008)
- [3] Pfleeger, Shari Lawrence, and Joanne M. Atlee. 2009. Software Engineering: Theory and Practice: Prentice Hall, ISBN 0136061699
- [4] Pressman, Roger S. 2010. Software Engineering : A Practitioner's Approach. 7th ed. New York: McGraw-Hill Higher Education, ISBN 9780073375977
- [5] Sommerville, Ian. 2010. Software Engineering. 9th ed, International Computer Science Series. Harlow: Addison-Wesley, ISBN 978-0137035151

Další literatura podle pokynů vedoucího diplomové práce.

Formální náležitosti a rozsah diplomové práce stanoví pokyny pro vypracování zveřejněné na webových stránkách fakulty.

Vedoucí diplomové práce: **Ing. Svatopluk Štolfa, Ph.D.**

Datum zadání: 01.09.2015

Datum odevzdání: 28.04.2017



doc. Dr. Ing. Eduard Sojka
vedoucí katedry

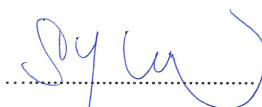


prof. RNDr. Václav Snášel, CSc.
děkan fakulty

Prehlásenie

Prehlasujem, že som túto diplomovú prácu vypracoval samostatne. Uviedol som všetky literárne pramene a publikácie, z ktorých som čerpal.

V Ostrave dňa 20.6.2017



.....

Podpis

PodĎakovanie

Ďakujem vedúcemu diplomovej práce Ing. Svatoplukovi Štolfovi, Ph.D., za nasmerovanie správnym smerom, za jeho rady, za delenie sa o vedomosti a nápady.

Abstrakt

Cieľom diplomovej práce je porozumieť oblasti process miningu a postupne prejsť všetkými definovanými úrovňami tejto problematiky – porozumenie, aplikácia, výskum, vývoj. Výsledkom práce bude návrh a vytvorenie novej metodiky verifikácie procesných modelov, ktorá bude implementovaná ako rozšírenie v nástroji ProM.

Kľúčové slová

Process mining, ProM, procesný model, log udalostí, XES, petriho siet'

Abstract

The objective of the Master Thesis is to understand the domain of the process mining and gradually passing through all defined levels of this issue – understanding, application, research, development. The result will be the design and implementation of a new methodology for process model verification, which will be implemented as an extension in the ProM tool.

Key Words

Process mining, ProM, process model, event log, XES, petri net

Obsah

Úvod.....	9
1. Process mining.....	10
1.1. Porozumenie hlavným oblastiam process miningu.....	10
1.2. Aplikácia process miningu za účelom vylepšovania procesov a procesných modelov.....	11
1.3. Expert výskumník.....	11
1.4. Developer.....	11
2. Oblasti process miningu.....	12
2.1. Objavovanie.....	12
2.2. Porovnávanie.....	12
2.3. Vylepšovanie.....	14
3. Rôzne uhly pohľadov na process mining.....	15
4. Aplikácia process miningu.....	16
4.1. Získavanie dát a práca s dátami z logov udalostí je najdôležitejší krok procesu.....	17
4.1.1. Prvá úroveň kvality dát.....	17
4.1.2. Druhá úroveň kvality dát.....	17
4.1.3. Tretia úroveň kvality dát.....	17
4.1.4. Štvrtá úroveň kvality dát.....	17
4.1.5. Piata úroveň kvality dát.....	17
4.2. Extrahovanie logov udalostí by malo byť riadené otázkami.....	18
4.2.1. Relevantné dáta v objednávkovom procese.....	18
4.3. Využívanie konštruktov postupností – konkurencia a rozhodovanie.....	18
4.4. Naväzovanie udalostí na prvky modelu.....	19
4.5. Model vystupuje ako úmyselná abstrakcia reality.....	19
4.6. Aplikácia process miningu je nepretržitý proces.....	21
5. Nástrahy a výzvy využívania techník process miningu.....	21
5.1. Vyhľadávanie, filtrovanie, triedenie, zlučovanie dát udalostí.....	21
5.2. Zavádzanie vhodných a zmysluplných hodnotení výsledkov.....	22
5.3. Dynamická podstata analyzovaných dát a procesov.....	22
5.4. Predsudky a predpoklady vyplývajúce z prezentácie výsledkov.....	23
5.5. Nepretržitý proces a priebežná podpora.....	23
5.6. Použitelnosť výsledkov process miningu aj pre neodborníkov.....	24
6. Algoritmy process miningu.....	25
6.1. Alpha algoritmus.....	25
6.1.1. Príklad – Vstupy.....	26
6.1.2. Príklad – Analýza vzťahov medzi aktivitami.....	27
6.1.3. Príklad – Procesný model.....	28
6.1.5. Zhrnutie.....	28
6.2. Evolučné algoritmy.....	29
7. Vyhodnocovanie výsledkov process miningu.....	30
7.1. Kritéria kvality.....	31
7.1.1. Presnosť.....	31
7.1.2. Jednoduchosť.....	31

7.1.3. Vhodnosť.....	32
7.1.4. Všeobecnosť.....	32
7.2. Hodnotenie algoritmov.....	33
7.2.1. Alpha algoritmus.....	33
8. Zdroje dát.....	33
8.1. Analýza zdrojov.....	34
8.2. Proces, prípad, udalosť.....	36
8.3. Best practices extrakcie a vytvárania logov udalostí.....	38
8.3.1. Súvislosť.....	38
8.3.2. Granularita.....	38
8.3.3. Rozsah riešenej oblasti.....	39
8.3.4. Časová známka.....	39
8.3.5. Podproces.....	39
8.4. Štandardné formáty logov udalostí.....	39
8.4.1. MXML.....	39
8.4.2. XES.....	40
9. Real-time process mining.....	41
9.1. Pre mortem a post mortem.....	41
9.2. Procesné modely.....	42
9.2.1. Kartografické procesné modely.....	43
9.2.2. Audit.....	43
9.2.3. Navigačné procesné modely.....	43
9.3. Operačný support.....	44
9.3.1. Detekcia.....	44
9.3.2. Predikcia.....	44
9.3.3. Doporučenie.....	45
10. Vymedzenie uplatnenia proces miningu v biznis procese.....	45
11. Metodika a verifikácia procesných modelov.....	46
11.1. Problém.....	46
11.2. Vstupné dáta.....	48
11.2.1. Pôvodný procesný model.....	49
11.2.2. Otázky, na ktoré sa snažíme nájsť odpovede verifikáciou procesného modelu	49
11.2.3. Dáta, ktoré zachytávajú skutočný obraz reality v analyzovanej organizácii.....	50
11.3. Analýza a návrh.....	51
11.3.1. Skúmanie a experimenty existujúcich možností analýzy modelu.....	51
11.3.2. Spracovateľný zápis položených otázok.....	55
11.3.3. Popis postupu navrhovanej metodiky.....	56
11.3.4. Toolchain.....	57
11.4. Implementácia.....	58
11.5. Výsledky verifikácie s využitím implementovaného pluginu.....	59
11.5.1 Konvertovanie heatmapy do formátu XES.....	59
11.5.2 Spustenie implementovaného pluginu.....	60
11.5.3 Výsledok verifikácie.....	61
11.5.3 Záver a návrhy na zmenu.....	63
12. Záver.....	65

Použitá literatúra.....	66
Zoznam príloh.....	67
A. ProM with implemented extension.....	67
B. Verification inputs.....	67
C. Video.....	67

Zoznam obrázkov

Obrázok 1. Úrovně aplikácie postupov process miningu.....	13
Obrázok 2. Metodika objavovania preberá na vstupe log událostí a na výstupe vydá vytvorený procesný model	14
Obrázok 3. Metodika porovnávania preberá na vstupe log událostí a pôvodný procesný model. Výstupom je diagnostika stavu procesu.....	15
Obrázok 4. Metodika porovnávania preberá na vstupe log událostí a pôvodný procesný model. Výstupom je zdokonalený pôvodný model.....	17
Obrázok 5. Aplikácia process miningu z pohľadu okolia [6].....	21
Obrázok 6. Schéma evolučného algoritmu [1].....	32
Obrázok 7. Kritéria kvality výsledkov postupov process miningu.....	34
Obrázok 8. Schéma znázorňujúce hlavné fázy metodík process miningu [1].....	37
Obrázok 9. Vzorka logu událostí v prirodzenej forme.....	39
Obrázok 10. Vzorka logu událostí vo formáte XES.....	43
Obrázok 11. Schéma navigačného procesného modelu.....	44
Obrázok 12. Fázy biznis procesu, v ktorých je možné a vhodné uplatniť metodiky process miningu.....	47
Obrázok 13. Schéma procesného modelu - Objednávkový proces od vstupu do systému po odoslanie objednávkového formulára.....	50
Obrázok 14. Události zaznamenané prostredníctvom heatmapy. ID prípadu je vlastne jednoznačný identifikátor návštevníka.....	53
Obrázok 15. Schéma pôvodného modelu analyzovaného procesu.....	54
Obrázok 16. Schéma formy Petriho siete vygenerovanej na základe reálneho logu událostí analyzovaného procesu.....	54
Obrázok 17. Triviálny graf aktivít na základe reálneho logu událostí analyzovaného procesu.....	55
Obrázok 18. Matica vzťahov aktivít.....	55
Obrázok 19. Flower model vygenerovaný na základe analyzovaného logu událostí.....	56
Obrázok 20. Formálny zápis pravidiel v nástroji Protege.....	57
Obrázok 21. Schéma postupu spracovania analyzovaných dát.....	58
Obrázok 22. ProM.....	59
Obrázok 23. Schéma verifikácie prípadov zachytených v logu událostí.....	60
Obrázok 24. ProM. Analýza logu událostí - analýza početnosti událostí v rámci jednotlivých inštancií procesu.....	61
Obrázok 25. ProM. Analýza logu událostí - analýza výskytu událostí.....	62
Obrázok 26: ProM. Analýza logu událostí - analýza koncových událostí zachytených prípadov.....	62
Obrázok 27. ProM. Voľba implementovaného pluginu.....	62
Obrázok 28. ProM. Výsledok verifikácie 1.....	63
Obrázok 29. ProM. Výsledok verifikácie 2.....	63

Úvod

Z výsledkov aktuálnych výskumov vyplýva predpoveď, ktorá hovorí, že veľkosť digitálneho vesmíru bude každé dva roky minimálne dvojnásobná. Dáta nás začínajú obklopovať na každom kroku. Spracovanie, filtrovanie, aplikovanie poznatkov vyplývajúcich z týchto dát, je často veľmi problematický a neefektívny proces. Okrem nástrah spojených so statickou stránkou vznikajúcich dát, ktorou sa zaoberajú oblasti big data a data miningu, sa objavujú otázky a možnosti spojené s dynamickou stránkou týchto dát. A práve touto stránkou sa zaoberá Process mining. Process mining zahŕňa postupy, metodiky, algoritmy, nástroje, ktoré analyzujú a spracovávajú dáta za účelom skúmania dynamickej (procesnej) stránky týchto dát.

Prvá kapitola obsahuje definície oblastí process miningu a definície úrovni porozumenia a osvojenia si problematiky process miningu

Druhá kapitola sa podrobnejšie venuje hlavným oblastiam process miningu, popisuje spoločné znaky metodík spadajúcich do danej oblasti. Súčasťou popisu je aj vizuálne znázornenie hlavných princípov postupov danej oblasti.

Tretia kapitola popisuje niekoľko možných perspektív, ktoré je možné aplikovať pri analýze procesného modelu. V niektorých prípadoch je dôležitá perspektíva založená

Štvrtá kapitola sa už zaoberá priamej aplikácii postupov process miningu. Kapitola zachytáva postup od kroku získania potrebných dát až po krok tvorby procesného modelu.

V piatej kapitole sú popísané nástrahy a prekážky, ktoré vyplývajú z využívania postupov a metodík process miningu.

Šiesta a siedma kapitola popisujú konkrétny algoritmus process miningu a spôsoby vyhodnocovania kvality výsledkov používaných algoritmov

Získavanie dát, best practices, formáty logov udalostí, táto problematika je popísaná v kapitole osem.

Kapitola deväť popisuje využitie process miningu v súvislosti s real-time dátami. Kapitola popisuje využitie v oblasti navigačných systémov, predikcií, v oblasti operačného supportu.

Krátka kapitola desať sa zaoberá vymedzením aplikácie process miningu v biznis procese.

Posledná kapitola predstavuje riešenie konkrétneho problému s využitím postupov process miningu. Súčasťou kapitoly je aj popis návrhu a prezentácia výsledkov novej metodiky verifikácie procesného modelu proti realite zachytenej v logu udalostí.

1. Process mining

Process mining zahrňuje nástroje a techniky, pomocou ktorých je možné využiť veľké množstvo dát vo svete okolo nás. Pomocou týchto dát je možné aplikovať a vytvárať nové formy analýz zlepšovania procesov. Process mining zahrňuje nástroje na odhalenie skutočných procesov vyplývajúcich z dát vzniknutých na základe skutočných udalostí v systémoch, nástroje na detekciu odchýlok, nástroje na analýzy využívania a prípadného mrhania zdrojov. Process mining by mal byť nástrojom každého pracovníka s dátami. Vytvára prepojenie medzi tradičným modelom analýzy procesov – napríklad simulácie – a technikami, ktoré sú analýzou zamerané na dáta – data mining, strojové učenie. Aplikácia správnych techník dáva zmysluplné výsledky v oblasti zlepšovania procesov rôznych domén. Dostupnosť rozsiahlych logov udalostí v kombinácii s nástrojmi process miningu umožňuje organizáciám diagnostikovať skutočné problémy pomocou faktov a nie pomocou hádania a odhadov. Process mining ale nie je oblasť, ktorej je možné jednoducho plne porozumieť na všetkých úrovniach. Vo všeobecnosti je možné chápanie a používanie process miningu vyjadriť v štyroch úrovniach [1]:

- Úroveň 1 – Porozumenie hlavným oblastiam process miningu
- Úroveň 2 – Aplikácia process miningu za účelom vylepšovania procesov a procesných modelov
- Úroveň 3 – Expert výskumník
- Úroveň 4 – Developer

1.1. Porozumenie hlavným oblastiam process miningu

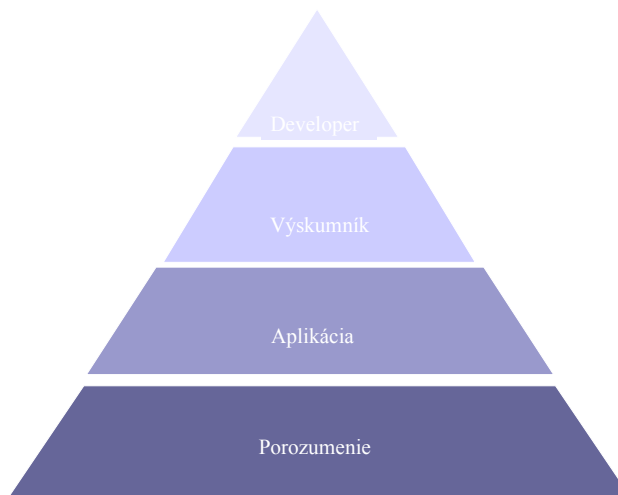
Postupy a techniky process miningu je možné rozdeliť do troch skupín – objavovanie, porovnávanie a vylepšovanie. V tejto kapitole popíšeme tieto oblasti [1] len krátko.

- Objavovanie – hlavnou myšlienkou postupov tejto oblasti je tvorba procesného modelu len s pomocou logov udalostí, bez pomoci ďalších dopĺňajúcich informácií. Jednou z techník tejto oblasti je Alfa algoritmus.
- Porovnávanie – vstupom techník tejto oblasti je log udalostí a existujúci procesný model, výsledkom je detekcia odchýlok a kontrola, či skutočný model (vyplývajúci z logov udalostí) odpovedá pôvodne navrhovanému modelu a opačne
- Vylepšovanie – aplikáciou techník tejto oblasti spracovávame informácie o aktuálnom skutočnom procese s cieľom rozvinúť a vylepšiť existujúci, pôvodný model. Príkladom môže byť napríklad rozšírenie modelu o výkonnostné dáta.

1.2. Aplikácia process miningu za účelom vylepšovania procesov a procesných modelov

V tejto úrovni sa už počíta so schopnosťou získavania zmysluplných dát spolu s ich previazaním na konkrétne problémy a otázky. Príkladom konkrétnych otázok sú:

- Aký je proces, podľa ktorého používateľa skutočne postupujú?
- Nachádzajú sa v procese miesta, v ktorých dochádza k zahlteniu?
- Aké faktory vplývajú na vznik zahltenia?
- Aké faktory vplývajú na vznik odchýlok a problémov v postupe?
- Je možné predikovať vznik týchto problémov?
- Aké možné protiopatrenia vyplývajú zo skutočných postupov?



Obrázok 1. Úrovne aplikácie postupov process miningu

1.3. Expert výskumník

Dosiahnutie tejto úrovne si už vyžaduje hlbšie pochopenie analytických techník a algoritmov process miningu, povedomie o technikách analýzy výkonu, problematika rozkladu existujúcich modelov, rozdeľovanie, transformácia a skladanie logov udalostí, zvažovanie kvality dostupných dát a ďalšie.

1.4. Developer

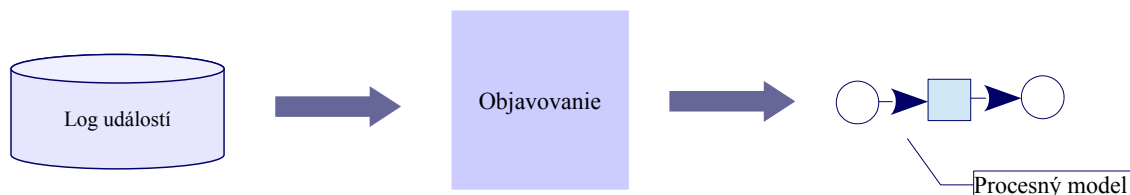
Vrcholom porozumenia oblasti PM sú schopnosti navrhovať nové techniky process miningu a ich následná implementácia, vytváranie nových nástrojov alebo rozširovanie existujúcich nástrojov, porovnávanie výsledkov nových techník s výsledkami existujúcich techník.

Ovládnutie úrovne 1 a úrovne 2 nám umožňuje používať nástroje a vylepšovať procesy. Úroveň 3 a úroveň 4 si už ale vyžadujú aj pochopenie, čo a ako sa vlastne pri analýze deje a ako tieto nástroje pracujú.

2. Oblasti process miningu

2.1. Objavovanie

Hlavnou myšlienkou techník, ktoré spadajú do tejto oblasti je tvorba, vytváranie, prezentácia. Nezáleží na spôsobe prezentácie ani na použitých vizualizačných nástrojoch. Jediným požiadavkom je, že postupy z tejto oblasti bude možno označiť ako tvoriace.



Obrázok 2. Metodika objavovania preberá na vstupe log udalostí a na výstupe vydá vytvorený procesný model

Organizácie používajú rôzne postupy na odbavenie scenárov vyplývajúcich zo splnenia požiadavkov na organizáciu. V niektorých prípadoch, alebo len v niektorých aktivitách, môžu byť tieto kroky scenára vynucované používaným informačným systémom. Vo väčšine sa ale stretávame s prípadmi, kedy zavedené používané postupy nie sú formálne a nie sú vhodne popísané. V tomto prípade neexistuje takmer žiadna možnosť nejakej analýzy a vylepšovania nastavených postupov a vytvára sa priestor pre vznik aktivít, ktoré vo výsledku môžu viesť k neefektívnemu spotrebovaniu zdrojov a nárastu nákladov. V horšom prípade môžu úplne zabrániť napĺňaniu podstaty organizácie – odbavovanie požiadavkov z okolia. Tieto prípady sú vhodné na zavedenie procesných modelov založených na dostupných logoch udalostí. Logy udalostí môžu byť data z rôznych zdrojov – ERP systém, transakčné logy, error logy, a podobne. Získané modely môžu slúžiť v rôznych situáciách, pri diskusii s vedením umožní vizualizovaný proces dosiahnutie jednotného pohľadu na vec, pri pokusoch o vylepšovanie a zefektívňovanie aktuálnych postupov, na formálne popísanie postupov pre budúcich členov organizácie.

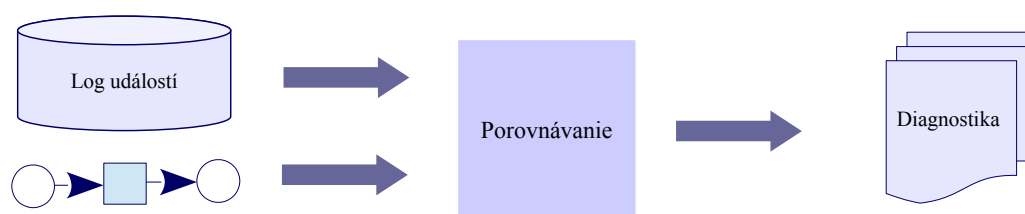
2.2. Porovnávanie

V niektorých prípadoch nemusíme pri pokusoch o tvorbu formálneho popisu procesných postupov začínať od nuly len s využitím logov udalostí. V prípade, že v skúmanej oblasti už existujú nejaké popísané modely, je možné pri aplikovaní process miningu využiť techniky zo skupiny porovnávaní.

Existujúci model mohol byť vytvorený ručne ešte vo fáze návrhu organizácie alebo mohol byť

založený v priebehu fungovania organizácie na základe hrubých popisov používaných postupov. Techniky porovnávania porovnávajú tieto existujúce procesné modely so skutočnými aktivitami, ktoré sú zachytené v logoch udalostí. Pri analýze sa zvyčajne postupuje postupným vytváraním väzieb medzi udalosťami z logu a aktivitami z dostupného procesného modelu. Prostredníctvom tohoto postupu je možné sledovať rozdiely medzi skutočne uskutočneným sa správaním a správaním, ktoré je zachytené v pôvodnom modeli. Príkladom časti výsledkov môže byť zistenie, že v danom modeli vytvorenom ešte vo fáze návrhu organizácie je uskutočniteľných 80% scenárov, ktoré vyplývajú z logu udalostí. Výsledky porovnávania je možné použiť hlavne na určenie kvality už zdokumentovaných modelov, do akej miery odpovedajú realite, ďalej k opísaniu scenárov, v ktorých dochádza k nežiadúcim odchýlkam a porozumieť tomu, čo je za vznikom týchto odchýlok. Ak sú v logu obsiahnuté aj dopĺňajúce informácie ako časové známky, informácie o zdrojoch alebo aktéroch, je následne možné presne určiť faktory, ktoré majú vplyv na vznik odchýlok. Metodiky porovnávania sa využívajú aj v prípade auditov, pri ktorých je nutné aby dokumentovaný stav presne odpovedal realite. A samozrejme sú techniky porovnávania vždy prvým krokom k ďalším návrhom na rozšírenie a zdokonalenie modelu.

Výsledky metodík zo skupiny objavovania majú jednu spoločnú charakteristiku, ktorá negatívne ovplyvňuje kvalitu získaných výsledkov. Na jednej strane všetky procesné modely vzniknuté pomocou týchto postupov znázorňujú len časť možného správania. Ani podrobné logy nemusia totiž zachytávať všetky možnosti systému – napríklad tie, ktoré ešte neboli využité. Toto zistenie by ale vo fáze objavovania nebolo možné pretože by neexistoval model, s ktorým by sme porovnali model vzniknutý čisto len z logov udalostí. Na druhej strane logy obsahujú úplné dáta o správaní užívateľov. Z pohľadu riešiteľa nás ale zaujíma pravdepodobne len určitá časť aktivít a preto je nutné tento nepotrebný šum odfiltrovať. Určenie, ktoré aktivity je možné považovať za šum nemusí byť triviálne – jednou z techník je odfiltrovať tie aktivity, ktorých výskyt je menší ako 80%. Aj v tomto prípade ale riskujeme strátu hodnotných a dôležitých aktivít, ktoré síce nemusia nastávať často ale z hľadiska organizácie alebo riešenej domény môžu byť veľmi dôležité. Metodiky porovnávania znižujú toto riziko práve využitím pôvodného modelu, na ktorého aktivity sú postupne naviazané udalosti z logu a postupne sa eliminuje pravdepodobnosť, že by sme z logu odfiltrovali udalosti, ktoré musia mať v modeli zastúpenie. Práve možnosti eliminácie šumu sú jednou z charakteristík, podľa ktorej môžeme hodnotiť algoritmy porovnávania.



Obrázok 3. Metodika porovnávania preberá na vstupe log udalostí a pôvodný procesný model. Výstupom je diagnostika stavu procesu.

Ďalšou charakteristikou hodnotenia kvality je predpokladanie nekompletnosti, kedy algoritmus počíta s tým, že v logoch udalostí nemusia byť zachytené úplne všetky informácie. Algoritmy, ktoré

predpokladajú prílišnú úroveň úplnosti logov vytvárajú model s prílišnou presnosťou. U takého modelu dochádza k vynechaniu aktivít, ku ktorým síce ešte nedošlo a nie sú teda súčasťou zaznamenaných akcií v logoch udalostí, sú to ale aktivity, ktoré majú v našej oblasti vysokú prioritu. A opačne ak algoritmus predpokladá s vysokou úrovňou nekompletnosti logu a vytvára v modeli príliš veľa priestoru na ďalšie aktivity, vzniká model s nízkou presnosťou pretože v tomto modeli je uskutočniteľných veľa scenárov, ku ktorým ale v realite nedochádza. Poslednou výraznou charakteristikou algoritmov zo skupiny porovnávania je schopnosť vysporiadať sa z prekážkami vyplývajúcich z prezentácie modelu. Niektoré algoritmy nedokážu vizuálne znázorniť hierarchiu alebo paralelismus. Aj tieto ukazatele môžu byť rozhodujúce pri voľbe vhodného algoritmu, pomocou ktorého sa pokúsime zodpovedať položené otázky. Algoritmy porovnávania využívajú v základe tri prístupy vykonávania porovnávania:

- Vytvorenie abstrakcie správania sa vyplývajúceho z dostupného logu a abstrakcie skúmaného procesného modelu – následne je možné sledovať vzťahy v týchto modeloch, napríklad ak v prvom modeli nasleduje vždy za udalosťou X udalosť Y a nikdy inak a v druhom modeli je tento vzťah porušený, je možné definovať odchýlku vo vzťahu postupnosti
- Simulovanie behu udalostí z logu nad analyzovaným modelom – metodika vyhodnocuje počet scenárov zachytených v logu, ktoré po spustení v počiatočnom bode modelu úspešne dosiahnu koncový bod modelu. Nevýhodou môže byť fakt, že táto metodika rovnako ohodnotí model, v ktorom scenár nie je možné ani spustiť a model, v ktorom scenár takmer dosiahne koncový bod. V tomto prípade je možné postupovať napríklad povolením všetkých hrán aj v prípade chýbajúcich tokenov a následne je možné do hodnotenia modelu aj započítať, koľko tokenov muselo byť umelo pridaných aby simulácia scenára dosiahla koncového bodu
- Určenie optimálnej zhody medzi aktivitou z logu a podobnou aktivitou z modelu. Úroveň perfektnej zhody je dosiahnutá vo chvíli keď všetky pohyby z logu sú pozorovateľné aj v modeli. Za nezhodu je považovaný stav, ak jeden alebo viacero pohybov zachytených v logu nie je možné odsledovať v analyzovanom procesnom modeli.

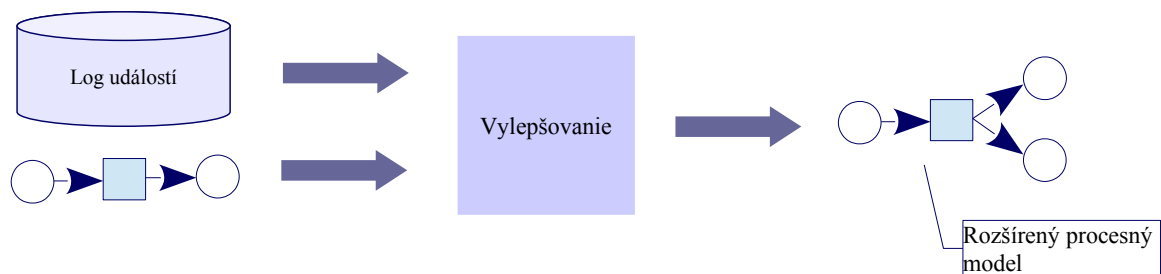
Na výsledky algoritmov porovnávania sa môžeme pozeráť z dvoch pohľadov. Prvý pohľad spočíva v myšlienke, že model nezachytáva reálne správanie sa a teda nesprávny a nedôveryhodný je spracovávaný model. Druhý pohľad spočíva v tvrdení, že udalosti zachytené v logu neodpovedajú požadovanej a plánovanej realite a teda, že nesprávny a chybný je aktuálne aplikovaný postup a používanie systému. Prvý pohľad je používaný v prípade, že cieľom process miningu je vytvorenie popisného procesného modelu, ktorý má zachytávať realitu alebo bude využívaný na tvorbu predikcií. Druhý pohľad je aplikovaný v prípade, že cieľom snaženia je vytvorenie modelu, na základe ktorého má dôjsť k úprave reality a teda k zmenám v aktuálne využívaných postupoch.

2.3. Vylepšovanie

Prostredníctvom techník a postupov skupiny vylepšovania je možné rozširovať a zdokonaľovať existujúce

modely. Dochádza k vylepšovaniu hlavne tých modelov, ktoré vznikli ešte vo fáze návrhu pred spustením daného systému.

Za vylepšenie modelu je možné okrem jeho korekcie pokladať aj napríklad doplnenie procesného modelu o presné údaje o využívaných zdrojoch alebo o časové údaje, ktoré sú zachytené v logoch udalostí. Predstavme si model objednávkového procesu od príchodu do systému až po dokončenie objednávky. Pôvodný model bude pravdepodobne obsahovať len jednotlivé aktivity a väzby medzi týmito aktivitami. Vylepšený model môže byť ale doplnený o časové údaje, pomocou ktorých bude možné z modelu vyčítať priemernú dobu, ktorú zákazníkovi zaberie dostať sa až do bodu dokončenia objednávky.



Obrázok 4. Metodika porovnávania preberá na vstupe log udalostí a pôvodný procesný model. Výstupom je zdokonalený pôvodný model.

Tu sa ponúkajú ďalšie možnosti analýz prípadných zahŕtení a hľadania miest, ktoré zákazníkovi pri nákupe zaberajú najviac času. Model doplnený o časové údaje môže byť využitý aj na predikciu zostávajúcej doby behu aktuálne spustených aktivít. Informácie o zdrojoch doplnené do procesného modelu môžu byť využité na definovanie nových rolí – skupiny entít, ktoré pravidelne vykonávajú určité aktivity. Vznikajú tu aj možnosti na popísanie rôznych sociálnych sietí.

Model doplnený o časové údaje môže byť využitý aj na predikciu zostávajúcej doby behu aktuálne spustených aktivít. Informácie o zdrojoch doplnené do procesného modelu môžu byť využité na definovanie nových rolí – skupiny entít, ktoré pravidelne vykonávajú určité aktivity.

3. Rôzne uhly pohľadov na process mining

Na aktivity a metodiky z oblasti process miningu je možné nahliadnuť z rôznych perspektív [1]. Zvolená perspektíva vyplýva z otázok, na ktoré sú aplikáciou postupov process miningu hľadané odpovede.

- kontrola priebehu – zameriava sa na kontrolu a sledovanie priebehu a postupnosti aktivít s cieľom nájsť a popísať všetky možné cesty. Výsledok môže byť reprezentovaný ako Petriho sieť, UML diagram, ...
- organizačná perspektíva – zameriava sa na dáta o skrytých zdrojoch v logu udalostí – napríklad

dáta o aktéroch (ľudia, role, iné systémy) – a ktoré sú zahrnuté do sledovanej aktivity. Výsledkom môže byť organizačná štruktúra znázorňujúca hierarchiu aktérov vyjadrenú pomocou rolí a zároveň môže zachytávať aj sociálne vzťahy a prepojenia

- perspektíva zameraná na scenáre – najčastejšie je scenár charakterizovaný prostredníctvom postupnosti krokov scenára alebo aktérmi, ktorí v scenári vystupujú. Okrem toho môžu byť ale scenáre charakterizované aj hodnotami elementov, ktoré v scenári vystupujú – napríklad počet objednaných položiek zachytených v objednávke
- časová perspektíva – zameriava sa na frekvenciu a časovanie jednotlivých udalostí. Ak máme v logoch udalostí k dispozícii časový údaj, je možné monitorovať vznik zahltenia, sledovať výkon a spotrebu zdrojov, predpovedať zostávajúci čas prebiehajúcich úloh

4. Aplikácia process miningu

Riešenie každej úlohy proces miningu začína plánovaním. Pred započatím riešenia musia byť najprv zozbierané všetky potrebné dáta udalostí, prípadné existujúce modely, musia byť určené hlavné všeobecnejšie ciele a otázky (konkretizácia otázok je už súčasťou aplikácie miningu). Zdrojmi týchto dát sú aktéri, management, iné systémy. Aby riešiteľ mohol správne rozhodnúť, ktorá časť z dostupných dát môže byť využitá, vyžaduje si táto fáza aspoň čiastočné porozumenie obsahu dostupných dát. A zároveň aby mohol riešiteľ pokladať tie správne otázky a aby mohol vlastne porozumieť výsledkom vyplývajúcich z miningu, je nutná znalosť a porozumenie analyzovanej oblasti. Prvé, automaticky generované (objavované) modely sú ešte veľmi úzko spojené s logmi udalostí, aj tieto už ale môžu poskytovať odpovede na niektoré otázky alebo poskytovať potrebné vstupy na vyvolanie návrhu na zmenu alebo prenasťavenie. Základný model môže byť využitý aj na odfiltrovanie veľmi výnimočných javov, ktoré môžu byť výsledkom nekompletných alebo celkovo nedokonalých dát a mohli by spôsobiť problémy v ďalších krokoch analýzy – je možné ich úplne odfiltrovať alebo prispôsobiť ďalší postup aj na tieto nepredpokladané stavy. Postupne je možné základný model rozširovať napríklad o zdroje alebo časové údaje. Výsledný model je možné využiť v operačnom supporte na ďalšie systémové zásahy, predikciu, odporúčenia.

Neodbornou a nedôslednou aplikáciou miningu môže dôjsť k vykonaniu určitých chýb, ktoré môžu mať za následok zníženie kvality a skreslenie výsledku. Existuje niekoľko definovaných pravidiel[8], pri uplatnení ktorých je možné výrazne znížiť pravdepodobnosť výskytu týchto chýb. Tieto pravidlá sú:

- Získavanie dát a práca s dátami z logov udalostí je najdôležitejší krok procesu
- Extrahovanie logov udalostí by malo byť riadené otázkami
- Využívanie konštruktov postupnosti - konkurencia a rozhodovanie
- Naväzovanie udalostí na prvky modelu
- Model vystupuje ako umýselná abstrakcia reality
- Aplikácia process miningu je nepretržitý proces

4.1. Získavanie dát a práca s dátami z logov udalostí je najdôležitejší krok procesu

Ako už bolo spomenuté, prvým krokom riešenia úlohy s využitím process miningu je zozbieranie dostupných dát, ktoré majú význam z hľadiska riešenej domény. Veľmi častou a opakovanou chybou je zameriavanie sa na zber len práve tých dát, ktoré sa nám javia ako ideálne za účelom dosiahnutia požadovaných výsledkov. Aplikovaním tohto nesprávneho postupu dochádza k nežiadúcemu vynechaniu práve tých dôležitých údajov, ktoré nám môžu odhaliť vznikajúce prekážky, ktoré nám aktuálne bránia k dosiahnutiu plánovaných cieľov a ktorých odhalenie má byť súčasťou výsledkov miningu. Zároveň je ale vhodné zbierať len tie údaje, ktoré majú pre nás význam z hľadiska riešenej úlohy. Potrebné dáta nemusia byť uložené v dedikovaných súboroch. Časti dát udalostí môžu byť obsiahnuté v databáze, v transakčných logoch, v emailových schránkach, v kartotékach, a podobne. Jednou z hlavných výziev fázy získavania dát je určenie kvality dostupných dát. Vo všeobecnosti je možné rozdeliť logy udalostí aj z hľadiska kvality do piatich úrovní [2].

4.1.1. Prvá úroveň kvality dát

Do úrovne najnižšej kvality zahrnujeme dáta, ktoré sú zaznamenané manuálne a teda nemusia vždy presne odpovedať realite. V prípade manuálneho záznamu je aj zvýšená pravdepodobnosť neúplnosti dát. Príkladom tejto úrovne sú papierové kartotéky pacientov[9].

4.1.2. Druhá úroveň kvality dát

Do druhej úrovne už patria dáta, ktoré sú zaznamenané síce automaticky, ale len ako vedľajší produkt nejakého systému – ich vytváranie nie je plánované. Pri zaznamenávaní týchto dát sa často nepostupuje podľa jasne určených pravidiel, čo sa bude zaznamenávať a čo nie. Z toho vyplýva, že najdôležitejšie aktivity nemusia byť v dátach tejto úrovne obsiahnuté. Príkladom informácií tejto úrovne sú error logy alebo podporné logy systémov riadenia.

4.1.3. Tretia úroveň kvality dát

Dáta tretej úrovne sú síce zaznamenávané automaticky, ani v tomto prípade ale nejde o systematické a cielené zaznamenávanie. Dostupné dáta nemusia byť úplné, na rozdiel od druhej skupiny ale tieto dáta odpovedajú realite a sú dôveryhodné – napríklad dáta z ERP systémov, transakčné logy.

4.1.4. Štvrtá úroveň kvality dát

Štvrtú skupinu tvoria logy, ktoré sú zaznamenávané plánovane, automaticky, systematicky a dôveryhodne. Dáta sú kompletne a odpovedajú skutočnosti. Hlavným rozdielom oproti dátam z predchádzajúcich úrovní je stav, kedy je z dát možné extrahovať úplné scenáre a aktivity. Príkladom sú heatmaps informačných systémov, ktoré zaznamenávajú úplné postupy užívateľov v systéme.

4.1.5. Piata úroveň kvality dát

Do najvyššej úrovne kvality dát zahrnujeme informácie, ktoré sú kompletne a dôveryhodné a ich

zaznamenávanie je jasne a presne definované. Udalosti a všetky ich doplňujúce atribúty majú jasnú štruktúru a sémantiku. Príkladom môžu byť logy udalostí systémov riadenia byznys procesov doplnené o sémantickú ontológiu.

4.2. Extrahovanie logov udalostí by malo byť riadené otázkami

Všetky aktivity process miningu – a teda aj získavanie dát – by mali byť riadené konkrétnymi otázkami. Extrakcia zmysluplných dát nie je možná bez toho, aby sme vedeli, ktoré dáta sú pre nás relevantné z hľadiska hľadaných odpovedí. Pre predstavu si môžeme vziať dáta z ERP systému, ktorý obsahuje tisíce tabuliek – napríklad celý bankový systém alebo systém nemocnice. Bez znalosti konkrétnych otázok nie je možné určiť tabuľky, ktoré môžu obsahovať žiadané logy udalostí.

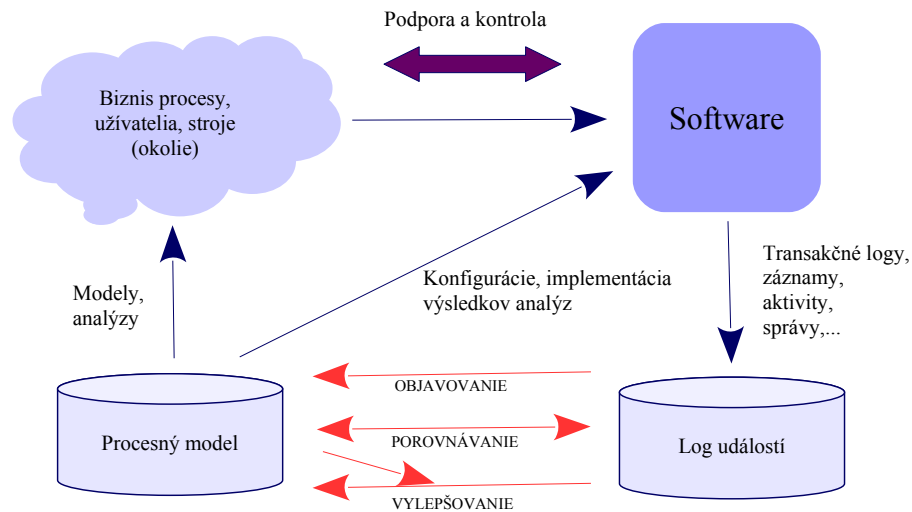
Pred započatím získavania dát by sme si mali zároveň zvoliť konkrétny scenár, ktorého proces chceme analyzovať. Taktiež táto voľba scenára by mala vyplývať z otázok, na ktoré hľadáme odpovede. Príkladom môže byť spracovanie zákazníckej objednávky.

4.2.1. Relevantné dáta v objednávkovom procese

Každá zákaznícka objednávka môže obsahovať viacero položiek pretože zákazník má možnosť objednať si v rámci jednej objednávky niekoľko rôznych produktov v rôznych konfiguráciách. Jedna zákaznícka objednávka môže byť rozdelená na niekoľko zásielok zákazníkovi – a teda aj rozdelená do niekoľkých nákladov dopravcu. Zároveň jeden náklad dopravcu môže obsahovať položky niekoľkých objednávok. Je vidieť, že medzi nákladom dopravcu a objednávkami vzniká väzba M:N, zároveň je vidieť, že tu ale vystupuje aj väzba 1:N medzi objednávkou a položkami objednávky. V týchto dátach objednávok, položiek, nákladov je možné sledovať niekoľko možných procesných modelov týkajúcich sa objednávkového procesu a logistiky. Ak dodržiavame doporučované postupy aplikácie process miningu a prvým krokom bolo plánovanie a polozenie správnych otázok, výrazne znižujeme riziko zvolenia dát, ktoré nie sú relevantné pre naše potreby. Zároveň nám toto uvedomenie môže uľahčiť rozhodovanie sa medzi skúmaním cyklu objednávky alebo cyklu prepravy zásielky.

4.3. Využívanie konštruktov postupností – konkurencia a rozhodovanie

V súčasnosti máme k dispozícii mnoho jazykov určených na tvorbu procesných modelov – BPMN, EPC, Petriho siete. Tieto jazyky poskytujú prvky, z ktorých je možné model vyhotoviť. Príkladom elementárnych prvkov sú miesta, prechody, hrany, príkladom zložených konštruktov určených na popis postupností sú sekvencia, paralelismus, rozhodovanie, cyklus. Tieto prvky sú podporované väčšinou z techník process miningu. Je veľmi dôležité kontrolovať, či aktuálne zvolený nástroj nebráni svojimi obmedzenými možnosťami v modeli plne znázorniť skutočne prebiehajúce procesy so všetkými relevantnými návaznosťami.



Obrázok 5. Aplikácia process miningu z pohľadu okolia [6]

4.4. Naväzovanie udalostí na prvky modelu

Ako už bolo spomenuté, aplikácia process miningu nezahŕňa len postupnosť udalostí. Obidve hlavné oblasti miningu – porovnávanie a vylepšovanie – zahŕňajú vo výraznej miere aj elementy v modeli a skúmajú vzťahy medzi týmito elementami a udalosťami v logu. Znalosť a analýza týchto vzťahov je potrebná predovšetkým v momente, kedy chceme v našom modeli nasimulovať priebeh konkrétnej sekvencie alebo spustiť určitú udalosť. Práve tieto simulácie a spúšťania udalostí nad naším modelom môžu odhaliť prípadné odchýlky a rozdiely medzi logom udalostí a modelom. Príkladom výsledku môže byť zistenie, že niektoré udalosti z logu nemôžu v našom modeli nikdy nastať – príčinou tohoto stavu ale môže byť nedokonalá voľba použitých dát pri tvorbe daného modelu. Ďalším príkladom výsledku, s ktorým sa môžeme pri navazovaní väzieb stretnúť je zistenie, že niektoré udalosti naopak v skutočnosti nastali ale sú vylúčené naším modelom – príčinou môžu byť možné trasy (vrátka) v aplikácii, ktoré sme nepredpokladali alebo neboli v prvotných fázach tvorby aplikácie dôležité ale môžu byť zdrojom problémov v systéme v budúcnosti.

Pri spracovaní dát môžu vznikať aj nejasnosti ohľadom toho, ku ktorej aktivite niektorá z udalostí skutočne patrí. Môže sa totiž zdať, že nejaká jedna konkrétna udalosť patrí k viacerým aktivitám alebo, že nepatrí k žiadnej aktivite. Takéto nejasnosti je nutné vyriešiť hneď na začiatku, v opačnom prípade riskujeme skreslenie výsledkov kedy udalosti jednej aktivity spracovávame ako udalosti viacerých aktivít a podobne. Celková interpretácia výsledkov teda nebude správna. V prípade, že je súčasťou dát z logu aj nejaká časová známka, je možné pomocou tejto informácie určovať dĺžku trvania určitej aktivity v závislosti na iných aktivitách.

4.5. Model vystupuje ako úmyselná abstrakcia reality

Každý model, ktorý je dôveryhodne vytvorený podľa kvalitných logov udalostí, znázorňuje určitý pohľad

na realitu. Logy udalostí môžu obsahovať dáta, pomocou ktorých je možné znázorniť niekoľko rôzne využiteľných pohľadov na realitu. Rôzni riešitelia aplikujú často rôzne techniky a postupy za účelom získania rôznych pohľadov. Model, ako výsledok spracovania logov udalostí, je možné v tomto prípade chápať ako nejaký druh mapy určitej oblasti. A podobne ako pre jednu geografickú oblasť je možné vytvoriť množstvo rôznych máp – mapa ciest, turistická mapa, mapa nerastných surovín, tak aj pre jednu oblasť zachytenú v procesných logoch je možné vytvoriť niekoľko modelov zachytávajúcich iné pohľady na túto oblasť – model zachytávajúci postupnosť aktivít, model zachytávajúci využiteľnosť zdrojov, model zachytávajúci role a aktérov, model zachytávajúci časovú perspektívu. Všetky tieto modely/mapy ale popisujú rovnakú časť reality. Každý model by mal zachytávať práve ten pohľad, ktorý má význam pre daného riešiteľa, alebo význam pre zadávateľa analýzy a tvorby modelu. Napríklad pre manažéra je dôležitý model, ktorý zachytáva realitu z pohľadu nákladov a príjmov a aj tento pohľad je možné vyjadriť na rôznych úrovniach, takže manažérovi môže postačovať len hrubý model. Analytik už ale vyžaduje vytvorenie podrobnejšieho modelu aby bolo možné sledovať prípadné odchýlky od plánovaného stavu. Jednotlivé zúčastnené strany riešiteľov a zadávateľov môžu teda vyžadovať pohľad na proces v rôznych úrovniach. Príkladom týchto úrovní sú:

- úroveň operačná – v tejto úrovni sú dôležité rozhodnutia, ktoré majú okamžitý dopad a sú založené na „čerstvých“ dátach práve bežiacich úloh
- úroveň taktická – do tejto úrovne vstupujú rozhodnutia so stredne dlhodobým dopadom. Využívajú sa aktuálne a nedávne dáta. Príkladom môže byť rozhodnutie manažéra na dočasné zníženie výdavkov na základe modelu znázorňujúceho výsledky z analyzovaných dát z aktuálnych a nedávnych príjmov
- úroveň strategická – rozhodnutia s dlhodobým dopadom – sú výsledkom analýzy dát zozbieraných za dlhšie časové obdobie

ďalšou z podobností, ktoré je možné pozorovať medzi tvorbou geografických máp a tvorbou modelov, je vizuálne odlišenie elementov napríklad podľa dôležitosti. V geografických mapách nemajú všetky mestá rovnakú veľkosť – záleží na rozlohe alebo počtu obyvateľov mesta. V procesnom modeli je možné meniť veľkosti prvkov aktivít napríklad podľa frekvenciu opakovania sa aktivity alebo podľa nákladov na priebeh aktivity. Rovnako význam väzby medzi jednotlivými aktivitami je možné znázorňovať rozličnými hrúbkami hrán a farbou hrany môžeme naznačiť miesto možného vzniku zahltienia.

Podobne ako pre jednu geografickú oblasť je možné vytvoriť množstvo rôznych máp – mapa ciest, turistická mapa, mapa nerastných surovín, tak aj pre jednu oblasť zachytenú v procesných logoch je možné vytvoriť niekoľko modelov zachytávajúcich iné pohľady na túto oblasť – model zachytávajúci postupnosť aktivít, model zachytávajúci využiteľnosť zdrojov.

Z popísaných možností a prípadov je jasné, že už na začiatku je nutné zahrnúť do plánovania všetky tieto varianty a vykonať rozhodnutia o spôsobe prezentácie modelu za účelom kvalitnejších a jednoduchších pochopiteľnejších výsledkov.

4.6. Aplikácia process miningu je nepretržitý proces

Správnou aplikáciou process miningu je možné dosiahnuť vytvorenie modelu, ktorý je založený na skutočne vykonaných akciách zaznamenaných v logoch udalostí. Zdroje týchto logov udalostí ale generujú zvyčajne dáta neustále – a to teda aj v momente analýzy a tvorby nášho modelu. Procesné modely zachytávajú dynamickosť/vývoj v určitom čase a z toho dôvodu sa proces mining snaží pozeráť na model a jeho analýzu ako na priebežný, nepretržitý proces a nie ako na jednorázový úkon. Naším cieľom by totiž nemala byť jednorázová tvorba modelu ale neustále dopĺňanie modelov o zmeny, ku ktorým v znázornenej realite dochádza neustále. Aj v tomto prípade je možné sledovať určitú podobnosť s klasickými mapami, napríklad dopravné mapy alebo mapy turistických atrakcií. V týchto mapách dochádza k zmenám často aj priamo vo chvíli, keď ich sleduje užívateľ – mapa dopravnej situácie v navigácii auta. Ak máme priebežne k dispozícii aktuálne logy udalostí, je možné rovnakým spôsobom aktualizovať aj vytvorené procesné modely. A rovnako ako môžeme sledovať zápchy v mape dopravnej situácie, je možné v prípade priebežnej aktualizácie modelu sledovať aj vznik zahľtenia a vďaka tomu predikovať a zobrazovať informácie o predpokladaných časoch dokončenia aktivít. Po pochopení týchto príkladov je jasné, prečo je vhodné pozeráť sa na process mining ako na nepretržitý proces a využívať ho na poskytovanie priebežných informácií.

A rovnako ako môžeme sledovať zápchy v mape dopravnej situácie, je možné v prípade priebežnej aktualizácie modelu sledovať aj vznik zahľtenia a vďaka tomu predikovať a zobrazovať informácie o predpokladaných časoch dokončenia aktivít.

5. Nástrahy a výzvy využívania techník process miningu

Process mining môže byť veľmi užitočným nástrojom. Na jednej strane musíme v súčasnej dobe riešiť extrémny nárast objemu dát, na druhej strane musíme prekonávať problémy spojené s filtrovaním, triedením zoskupovaním vznikajúcich informácií a procesov aby sme mohli plniť požiadavky vyplývajúce z efektivity a z poskytovania služieb. K tomuto účelu je možné využiť práve process mining, je nutné ale myslieť aj na výzvy a nástrahy, s ktorými sa môžeme pri práci stretnúť. Keďže hlavnou myšlienkou miningu je analýza dynamickosti, aj nástrahy vyplývajúce z tejto analýzy nemusia byť trvácne a je pravdepodobné, že niektoré aktuálne nástrahy budú postupne eliminované pokrokom[1] a naopak je správne predpokladať vznik nových nástrah.

5.1. Vyhľadávanie, filtrovanie, triedenie, zlučovanie dát udalostí

Už samotná extrakcia dát, ktoré sú skutočne vhodné z hľadiska riešenej otázky a sú využiteľné technikami process miningu, si môže vyžadovať veľa času a úsilia. Pri tomto kroku môžeme naraziť na niekoľko prekážok.

V prípade, že dáta pochádzajú z rôznych zdrojov, je nutné tieto dáta zlúčiť. Tento úkon ale nemusí byť úplne triviálny hlavne v prípade, ak tieto zdroje používajú rozličné identifikátory – v jednom zdroji môže byť jednoznačným identifikátorom rodné číslo, v inom zdroji môže byť identifikátorom emailová adresa. Dáta z logov udalosti sú často naviazané aj na nejaký objekt alebo scenár. Napríklad odkazujú na produkt, objednávku, zásielku. Ak chceme vytvoriť a umožniť monitorovanie kompletného procesného modelu od príchodu zákazníka do eshopu až po odovzdanie zásielky zákazníkovi, musíme objaviť a popísať väzby medzi aktivitami rôznych systémov – eshop a systém logistiky dopravcu. Získané dáta nemusia obsahovať všetky potrebné informácie, nemusia vôbec dochádzať k logovaniu aktivít a stavov dôležitých objektov, ktoré do procesu vstupujú. Alebo nemusia obsahovať úplné časové údaje. Zároveň sa v zaznamenaných údajoch môžu vyskytovať informácie rôznej granularity. Príkladom môžu byť logy dodávateľa, ktoré môžu obsahovať napríklad len data o dodaní zásielky konečnému zákazníkovi. Môžu ale obsahovať aj data zaznamenávajúce proces zásobovania iných eshopov a veľkoodberateľov. Pri získavaní dát je nutné počítať aj faktom, že zdroje môžu poskytovať dáta rôzneho rozsahu. Niektoré zdroje môžu poskytovať dáta takého rozsahu, že je veľmi náročné s týmito dátami efektívne pracovať. Iné zdroje môžu zasa poskytovať tak málo dát, že nie je na ich základe možné vyvodiť spoľahlivé závery.

Hlavným cieľom pri získavaní a organizovaní dát je dosiahnuť na dáta najvyššej kvality. Pri prekonaní uvedených nástrah a prekážok sa môže kvalita dát značne zvýšiť.

5.2. Zavádzanie vhodných a zmysluplných hodnotení výsledkov

Jedným z najväčších problémov, s ktorými sa process mining stretáva, je absencia stopercentne kvalitných spôsobov hodnotenia výsledkov [2]. Aj napriek tomu, že v súčasnosti poznáme niekoľko spôsobov vytvárania modelov na základe logov udalostí, stále nie sú k dispozícii techniky na dôveryhodné vyhodnotenie kvality dosiahnutých výsledkov. V súčasnosti je možné techniky hodnotiť len podľa rozsahu možností tejto techniky, aby bolo ale zaručené spoľahlivé hodnotenie, je nutné neustále vytvárať a vyhodnocovať nové techniky merania kvality. Cieľom je získať techniky, ktoré budú pozostávať z ukázkových dát a presných kritérií kvality, podľa ktorých bude možné výsledok vyhodnotiť.

Novo vznikajúce metodiky sa pri hodnotení zameriavajú predovšetkým na tri veličiny [2] – jednoduchosť, presnosť, všeobecnosť. Pri určovaní kvality totiž nie je vhodné hodnotiť len výsledky nad reálnymi dátami ale je správne hodnotiť výsledky vzniknuté nad dátami umelo vytvorenými. Vďaka tomu je zaručené, že napríklad kladne hodnotené modely, zachytávajú okrem len skutočne vyskytnutých sa udalostí aj tie udalosti, ktoré môžu nastať ale ešte k nim nedošlo. Zároveň by sa kladne hodnotené modely mali vysporiadať aj s prípadmi, kedy logy obsahujú veľa nepotrebných a neužitočných informácií (šum) a model by ich mal alebo ignorovať alebo im prikladať menšiu prioritu ako informáciám, ktoré majú pre nás význam z pohľadu riešených otázok.

5.3. Dynamická podstata analyzovaných dát a procesov

Predstavme si nákupný proces, kedy v tej istej organizácii môže byť ten istý proces vykonávaný inak v

závislosti na okolí. Napríklad počas roka môže byť proces vytvorenia objednávky, odbavenia a potvrdenia objednávky, balenie zásielky, odoslanie zásielky zložený z aktivít, ktoré sú vykonávané výhradne sekvenčne a žiadna z aktivít nie je započatá pred riadnym dokončením predchádzajúcej aktivity. V predvianočnom období, kedy je počet odbavených objednávok najvyšší, môže ale dôjsť k zmene v postupe odbavovania objednávky a niektoré aktivity môžu byť vykonávané paralelne. Napríklad balenie zásielky môže byť spustené ešte pred úplným odbavením a potvrdením objednávky. Zároveň môže v tomto období dôjsť aj k navýšeniu počtu zamestnancov v organizácii, čo taktiež môže vyústiť k úpravám vo vykonávaných postupoch.

Vytváraný a analyzovaný procesný model sa teda môže meniť a vyvíjať neustále. Aj tieto zmeny je nutné zahrnúť do nášho chápania a aplikovania techník process miningu. V prípade, že tieto zmeny nie sú súčasťou našej analýzy, vzniká tak priestor na negatívny dopad týchto zmien na finálne výsledky. Výskyt týchto problematických zmien môžeme znižovať rozdeľovaním logov udalostí na menšie časti a riešiť analýzu a tvorbu procesov na základe týchto menších celkov.

5.4. Predsudky a predpoklady vyplývajúce z prezentácie výsledkov

Aj keď sa to môže javiť ako nie úplne štandardný postup, v process miningu je dôležité aby sme sa vo fáze plánovania, pokladania otázok, objavovania úplne oprostili od plánovania spôsobu vizualizácie [1]. V opačnom prípade sa zvyšuje riziko neúplných výsledkov prezentovaných finálnou vizualizáciou spôsobené nezámerných vynechaním tých udalostí a podprocesov, s ktorými by mohol byť problém pri vizualizácii vo zvolenej technike prezentácie výsledkov. Voľba a zber spracovávaných dát by mala byť vedomá a cielená aktivita a nemala by podliehať preferovanému spôsobu prezentácie. Napríklad fakt, uprednostňovaná voľba prezentačnej techniky, ktorá neumožňuje znázornenie konkurencie, môže mať nežiadúci výrazný dopad na tvorbu modelu.

5.5. Nepretržitý proces a priebežná podpora

V súčasnosti dosiahlo ľudstvo stav, kedy má vďaka pokroku v technológiách konečne k dispozícii dostatočný výpočtový výkon, ktorý nám umožňuje spracovávať a analyzovať neustále narastajúce množstvo dát a udalostí a to už aj v momente ich výskytu.

Priebežné monitorovanie a kontrola každej udalosti by mala byť riadená tromi krokmi – detekcia, predikcia, odporúčanie. Ak v monitorovanom procese/scenári dochádza k odchýlke, je možné tento stav okamžite detekovať, vyhodnotiť, a v prípade potreby vygenerovať upozornenie. Neustály monitoring je vyžadovaný hlavne z dôvodu znižovania rizika vzniku prípadnej škody. Ak je odchýlka detekovaná ešte vo chvíli, keď k odchýlke došlo, je možné na základe vygenerovaného upozornenia vykonať potrebné kroky a zasiahnuť do behu scenára za účelom odstránenia odchýlky alebo zníženia rozsahu následkov. Zároveň historické data priebežného monitoringu sa ponúkajú na spracovanie s cieľom vytvárania predikcií a prediktívnych modelov, ktoré môžu byť zasa analyzované a porovnávané s modelmi bežiacich procesov, a výsledkom môže byť predikcia zostávajúceho času práve bežiacich aktivít. Z týchto odhadov a prediktívnych modelov môžu ďalšou analýzou vzniknúť návrhy na zmeny v systéme, ktoré môžu viesť k

znižovaniu potrebného času na vykonanie aktivít a na znižovanie nákladov vyplývajúcich z behu aktivít.

5.6. Použitelnosť výsledkov process miningu aj pre neodborníkov

Jednou z hlavných myšlienok process miningu je tvorba živých procesných modelov, ktoré by boli využívané denne. Na rozdiel od statických modelov, ktoré už len z podstaty dynamickosti procesu, sú skôr či neskôr odsúdené na založenie do archívu, dynamické modely môžu byť prostredníctvom spracovania neustále vznikajúcich nových dát využité na objavovanie nových nastávajúcich udalostí. Zmysluplné prepojenie vytvoreného procesného modelu a priebežne vznikajúcich nových logov udalostí nám umožní prezentovať do modelu aktuálny stav reality. Tento postup umožňuje užívateľom prácu s výsledkami process miningu na dennej báze.

Neustále aktualizované dáta, na základe ktorých je možné napríklad vykonávať aj okamžitú predikciu nasledujúcich udalostí alebo spotrebovaných zdrojov, môžu byť veľmi cenné. Takzvané real-time dáta ani nemusia byť pri nepretržitom spracovaní ani vždy riadne ukladané ale môžu byť okamžite spracované a analyzované už vo chvíli ich objavenia sa. Ide hlavne o prípady, kedy rýchlosť prezentácie informácií má oveľa vyššiu prioritu ako ich uloženie. Napríklad dáta obchodovania na burze, kedy uloženie aktuálneho stavu a cien ponuky a dopytu nie je tak dôležité ako zobrazenie týchto dát. Každá sekunda môže predstavovať rozdiel medzi ziskom alebo naopak veľkými stratami. Súčasne prípadné prediktívne modely, ktoré sú aktualizované nepretržite a predpovedajú vývoj na burze, môžu byť neoceniteľným nástrojom pri navyšovaní ziskov.

V process miningu je dôležité aby sme sa vo fáze plánovania, pokladania otázok a objavovania úplne oprostili od plánovania spôsobu vizualizácie – vzniká riziko neúplných výsledkov prezentovaných finálnou vizualizáciou spôsobené nezámerným vynechaním tých udalostí a podprocesov, s ktorými by mohol byť problém pri vizualizácii vo zvolenej technike.

ďalším príkladom môže byť dopravný navigačný systém, ktorý nás pomocou nepretržitého zberu a analýzy dát môže včasne informovať o prípadnom vzniku dopravnej zápch (zahľtenia). Nemusí vždy platiť, že dáta sa do aktualizovaného modelu dostanú vždy okamžite, záleží aj na vzniku prípadných zahľtení pri predávaní údajov a podobne. Je možné ale tvrdiť, že aj keď v minulosti slúžili predovšetkým na zachytávanie dát do nejakého úložiska, z ktorého následne mohol byť vytvorený nejaký statický model, v dnešnej dobe je možné vďaka technologickému pokroku dokonca aktualizovať modely ešte pred uložením spracovávaných dát.

Úlohou process miningu je aj zaistiť využívanie týchto dát a výsledkov analýz aj pre neodborníkov. Preto je vhodné pred užívateľom skryť komplexné algoritmy analýzy a tvorby procesných modelov a poskytnúť užívateľovi jednoduché rozhranie, ktoré mu umožní upraviť určité parametre nastavenia za účelom dosiahnutia odpovedí na položené otázky. Zároveň okrem použiteľného rozhrania nastavenia je nutné zaoberať sa aj zrozumiteľným a zmysluplným zobrazovaním prezentovaných výsledkov. Pri voľbe a implementovaní spôsobu zobrazenia výsledkov neodbornému užívateľovi musíme

myslieť aj na riziko, že tento užívateľ neporozumie prezentovaným informáciám a toto riziko sa snažíme eliminovať správnou variantou prezentácie. V opačnom prípade užívateľ výsledkom neporozumie a v horšom prípade získa mylnú predstavu o realite a neuvedomí si, že jeho chápanie výsledkov nie je správne.

6. Algoritmy process miningu

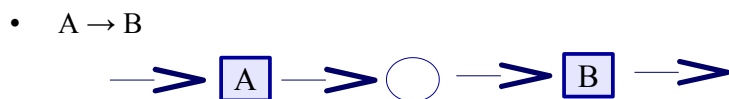
6.1. Alpha algoritmus

Jedným z prvých algoritmov využívaných na vytváranie procesných modelov bol alpha algoritmus. Tento algoritmus preberá na vstupe dáta zaznamenané v logu udalostí a na výstupe vyprodukuje Petriho sieť, v ktorej je možné spustiť jednotlivé scenáre vyplývajúce z analyzovaného logu.

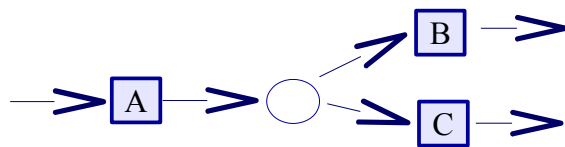
Algoritmus prechádza log udalostí a vyhľadáva v zaznamenaných aktivitách určité vzorce. Napríklad ak je aktivita A nasledovaná aktivitou B ale aktivita B nie je nikdy nasledovaná aktivitou A, je možné vydedukovať jednoduchú závislosť medzi A a B. Pri analyzovaní logu počas aplikácie alpha algoritmu vyhľadávame štyri možné vzorce, ktoré vyjadrujú nejaký vzťah v poradí jednotlivých aktivít:

- $A > B$: B je možné označiť za nasledovníka aktivity A ak existuje aspoň jedna trasa X, pre ktorú platí
 - $X = \langle x_1, x_2, x_3, \dots, x_n \rangle$ a $i \in \{1, \dots, n-1\}$, $x_i = A$ a $x_{i+1} = B$
- $A \rightarrow B$: medzi aktivitou A a aktivitou B je možné definovať kauzálnu závislosť (príčina / následok) ak platí
 - $A > B$ ale zároveň v celom logu platí $B \nexists A$
- $A \parallel B$: paralelnú závislosť medzi aktivitou A a aktivitou B (v sekvencii môžu mať rovnaké poradie) definujeme ak platí
 - $A > B$ ale zároveň v logu existuje aj $B > A$
- $A \# B$: medzi aktivitou A a aktivitou B neexistuje žiadna závislosť ak platí
 - $A \nexists B$ a zároveň pre celý log platí $B \nexists A$

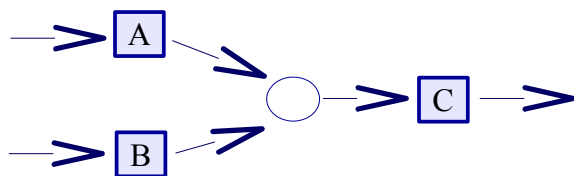
Formálne popísané vzťahy medzi aktivitami je následne možné jednoducho vyjadriť vizuálne prostredníctvom množiny konštruktov stavov a prechodov:



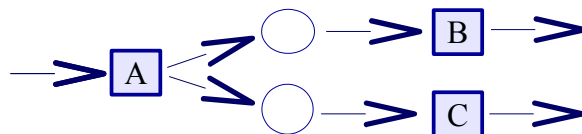
- XOR rozdelenie: $A \rightarrow B, A \rightarrow C, B \# C$



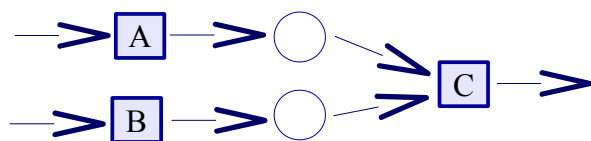
- XOR zlúčenie: $A \rightarrow C, B \rightarrow C, A \# B$



- AND rozdelenie: $A \rightarrow B, A \rightarrow C, B \parallel C$



- AND zlúčenie: $A \rightarrow C, B \rightarrow C, A \parallel B$



6.1.1. Príklad – Vstupy

Log udalostí obsahuje zachytené aktivity z procesu sledovania agendy platobných transakcií. Zaznamenané aktivity je možné zoskupiť do piatich najčastejších scenárov.

Trasa	Početnosť
AEDI	32
ADEI	4
AFI	21
AFGHI	16
AFHGI	19

Kód	Aktivita
A	Vstup do sekcie
D	Tisk dokladu
E	Zobrazenie zostatku
F	Vstup do histórie transakcií
G	Voľba podľa termínu
H	Voľba podľa typu
I	Výstup zo sekcie

6.1.2. Príklad – Analýza vzťahov medzi aktivitami

Priamy nasledovník
A > E H > I
A > D G > I
A > F E > I
A > E F > H
E > D F > G
D > I F > I
D > E

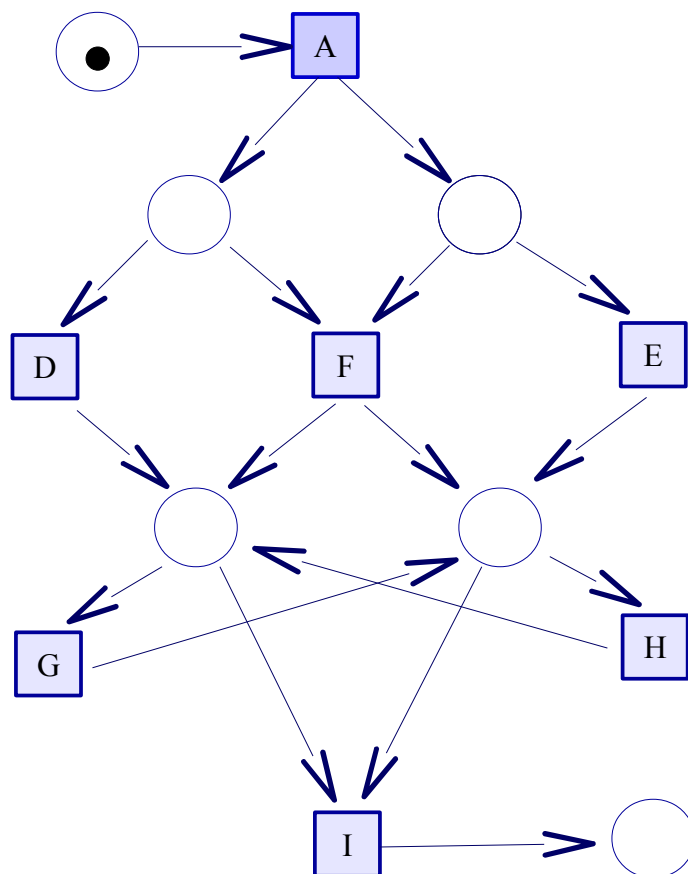
Kauzalita
A → D F → G
A → E F → H
A → F E → I
D → I G → I
F → I H → I

Paralelismus
E D
D D
G H
H G

Matica vzťahov

	A	D	E	F	G	H	I
A	#	→	→	→	#	#	#
D	←	#		#	#	#	→
E	←		#	#	#	#	→
F	←	#	#	#	→	→	→
G	#	#	#	←	#		→
H	#	#	#	←		#	→
I	#	←	←	←	←	←	#

6.1.3. Príklad – Procesný model



6.1.5. Zhrnutie

Ako je znázornené v predchádzajúcom príklade, alpha algoritmus vytvára procesný model len za pomoci logu udalostí. Táto metodika ale predpokladá, že analyzovaný log udalostí obsahuje všetky a kompletne možné scenáre, ktoré sú pre riešiteľa dôležité z hľadiska riešenia položených otázok process miningu. V skutočnosti ale nie je vhodné predpokladať úplnosť logu. Riešený systém môže obsahovať udalosti, ku ktorým ešte nedošlo a preto nebudú súčasťou logovaných informácií.

V dnešnej dobe poznáme ale niekoľko spoľahlivejších a dôveryhodnejších, zložitejších algoritmov aplikovateľných pri tvorbe modelu, ktoré poskytnú na základe poskytnutých dát kvalitnejší model ako alpha algoritmus. Nedostatkom tohto jednoduchého algoritmu je hlavne neschopnosť vyrovnáť sa s možným šumom v logu, neschopnosť spoľahlivo vyhodnotiť nekompletný scenár alebo spracovať a

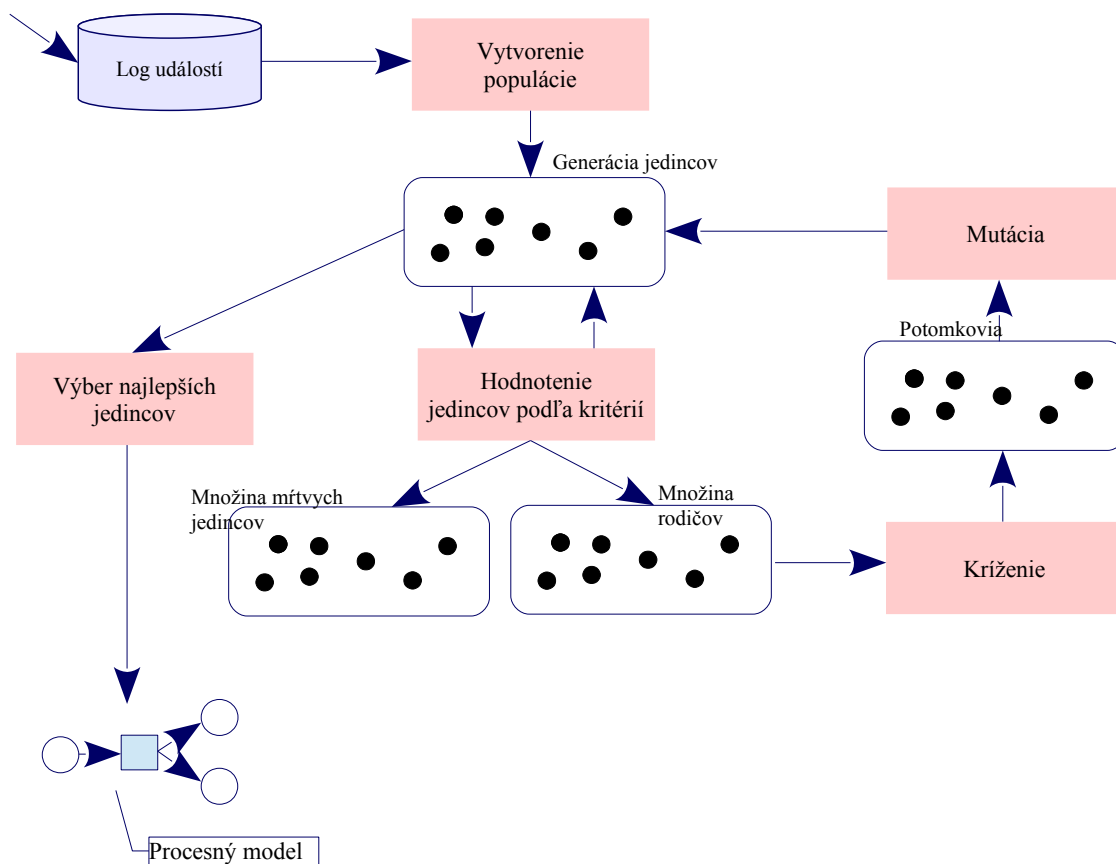
zachytiť v modeli zložitejšie konštrukcie. Z toho dôvodu sa alpha algoritmus v praxi uplatňuje predovšetkým v teoretickej rovine na tvorbu ukážkových modelov.

6.2. Evolučné algoritmy

Medzi algoritmy typu porovnávania je možné zaradiť aj evolučné algoritmy. Na rozdiel od algoritmov objavovania, ktoré vytvárajú procesný model priamo a deterministicky, evolučné algoritmy vytvárajú model iteratívne. Pomocou nedeterministického prístupu a vyhľadávaním nových alternatív prostredníctvom vnášania prvku náhodnosti, sa tieto algoritmy snažia nasimulovať process prirodzenej evolúcie. Evolučné algoritmy postupujú podľa štyroch hlavných krokov [3]:

1. Vytvorenie populácie
2. Výber najvhodnejších jedincov
3. Reprodukcia
4. Ukončenie

Cieľom akcií vykonávaných v kroku vytvorenia populácie je vytvorenie jedincov na základe dostupného logu. V tomto prípade je jedincom konkrétny procesný model, ktorý vzniká podľa aktivít zachytených v logu a výsledný model je tvorený úplne náhodne. Keďže vytvorený procesný model nedodrížiava postupnosti akcií z logov, môže v každej generácii vzniknúť aj tisícovka jedincov v závislosti len na počtu aktivít logu. Je pravdepodobné, že jedinci prvej generácie nebudú spĺňať kritérium presnosti, pretože aj keď tieto modely budú tvorené rovnakými aktivitami ako riadne procesy v systéme, správanie, ktoré budú znázorňovať bude od skutočne zachyteného správania sa v logu veľmi odlišné. Následkom náhodných akcií a veľkého počtu týchto jedincov sa postupne u náhodných modelov budú objavovať časti procesných modelov, ktoré už budú odpovedať správaniu sa zachytenému v logu. Výber týchto častí a teda aj výber konkrétnych jedincov už radíme do druhého kroku. V tomto kroku sa určuje kvalita jednotlivých náhodných modelov podľa ich vzťahu k logu. Keďže nie je známy univerzálny spôsob určovania kvality modelu, postupuje sa najčastejšie podľa otázky “Koľko trás vyjadrených v logu je možné nasimulovať v náhodnom procesnom modeli”. Po získaní jedincov najvyššej kvality nasleduje tretí krok, reprodukcia. Akcie tohoto kroku môžeme rozdeliť do dvoch fáz - mutácia a kríženie. Ak by sme pri reprodukcii nepristúpili aj k mutácii, všetky ďalšie vytvorené modely by boli zviazané možnosťami modelov prvej populácie.



Obrázok 6. Schéma evolučného algoritmu [1]

Vo fáze kríženia sa tvoria nové modely krížením rodičovských procesných modelov, nasleduje fáza mutácie, ktorá zabezpečí prídanie alebo odobratie nového prvku. Po dokončení generovania jedincov sa opakuje druhý krok, kedy sú zvolení jedinci s najsilnejšou presnosťou voči spracovávanému logu. Do kroku ukončenia vstupujeme až v momente, keď sú splnené požadované kritéria presnosti a vytvorený procesný model má dostatočnú kvalitu.

7. Vyhodnocovanie výsledkov process miningu

Úlohou aktivít process miningu nie je len vytvárať a zdokonaľovať procesné modely. Jednou z hlavných činností je aj vytváranie a definovanie nových, spoľahlivých techník a postupov na určovanie kvality algoritmov vytvárajúcich procesné modely.

V algoritmoch process miningu vystupuje napríklad na rozdiel od algoritmov klasického data miningu oveľa viac perspektív, ktoré je nutné zahrnúť do výsledného ohodnotenia výsledkov. Modelácia procesov je komplexný proces a v súčasnosti poznáme veľké množstvo postupov vytvárania modelov, ktorých výsledky sa môžu líšiť z hľadiska podrobnosti (modely znázorňujú len najčastejšie spustené aktivity) alebo napríklad z hľadiska prezentácie (môže byť náročné porovnať dva procesné modely

vytvorené dvoma rozdielnymi prezentačnými princípmi). Vo všeobecnosti definujeme štyri základné kritériá, ktorých úroveň splnenia určuje celkové hodnotenie analyzovaného procesného modelu [2]:

- presnosť
- jednoduchosť
- vhodnosť
- všeobecnosť

7.1. Kritéria kvality

7.1.1. Presnosť

Existuje niekoľko spôsobov, pomocou ktorých môžeme ohodnotiť presnosť modelu. Jedným z možných postupov je využitie množín dostupných aktivít.

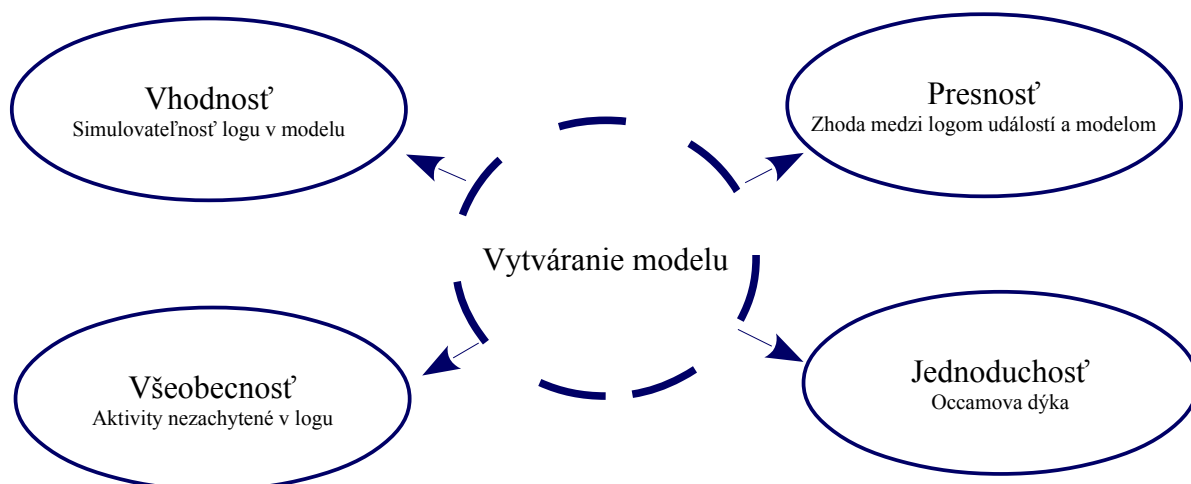
Je definovaná množina všetkých skupín udalostí z logu, množina stavov procesného modelu a množina aktivít, u ktorých predpokladáme väzbu na udalosť z logu. Ak sa predpokladá, že log bol pretvorený do procesného modelu postupným porovnávaním a pridávaním stavov do modelu počas prechádzania jednotlivých udalostí logu, je následne možné predpokladať, že každá udalosť má zastúpenie v modeli a existuje väzba medzi touto udalosťou a stavom modelu. Pri dodržaní tohto postupu získame vo výsledku model s vysokou presnosťou. V tomto prípade ale dochádza k problému, kedy vytvorený model umožňuje len presne to správanie sa, ako je zachytené v logu udalostí. Tento stav nemusí byť žiaduci, napríklad z dôvodu nekompletnosti dostupných dát. Vytvorený model by totiž znázorňoval len časť skutočne vykonávaných postupov. Na druhú stranu ale nie je vhodné aby výsledný model mal príliš nízku úroveň presnosti. Veľmi nepresný model totiž síce umožní vykonávanie všetkých scenárov zachytených v spracovávanom logu, umožní ale aj veľa ďalších možných scenárov, ktoré ale v skutočnosti nikdy nenastanú. Z toho teda vyplýva, že presnosť modelu vlastne vyjadruje pomer objemu scenárov zachytených v logu a objemu možných scenárov, ktoré sú umožnené vytvoreným modelom ale nedošlo k ich výskytu v organizácii a teda nie sú dostupné v logu udalostí. V skutočnosti ale vykonávame len odhady presností – takmer každý model môže mať totiž nekonečné množstvo možných scenárov, čo je spôsobené výskytom cyklusov v procesnom modeli.

7.1.2. Jednoduchosť

Hodnotenie podľa kritéria jednoduchosti vyjadruje, ako jednoducho je získaný a prezentovaný procesný model zrozumiteľný pre človeka. Pravidlo nazývané Occamova dýka tiež hovorí „Ak pre nejaký jav existuje viacero vysvetlení, je lepšie uprednostňovať to najmenej komplikované“. A práve toto pravidlo sa uplatňuje v prípade, že je cieľom dosiahnutia čo najvyššej úrovne jednoduchosti modelu.

Toto kritérium nemá priamu a rozhodujúcu väzbu na fakt, či model skutočne odpovedá reálnemu správaniu sa. V niektorých prípadoch je ale jedinou možnosťou zjednodušenia komplexného modelu úprava správania sa v modeli. Týmto zásahom ale dochádza k ovplyvňovaniu ostatných kritérií kvality

algoritmov process miningu. Jedným z hlavným faktorov určujúcim komplexnosť prezentovaného modelu je veľkosť modelu.



Obrázok 7. Kritéria kvality výsledkov postupov process miningu

7.1.3. Vhodnosť

Pomer medzi objemom scenárov, ktoré vyplývajú z logu udalostí a sú nasimulovateľné v analyzovanom procesnom modeli a objemom scenárov, ktoré nie sú spustiteľné v danom modeli, je vyjadrený kritériom vhodnosti. V súčasnosti je k dispozícii niekoľko techník, ktorých hlavnou náplňou je sledovanie tohto kvalitatívneho kritéria.

Súčasný postup sa líši napríklad v metodike ohodnotenia modelov, ktoré dosiahnú po spustení v modeli prijímajúci konečný stav a modelov, ktoré môžu ostať v aktívnom stave aj po kompletnom dokončení simulovaného scenára z logu. Niektoré postupy akceptujú aj tieto na pohľad nedokonalé procesné modely ako kvalitné. Ďalším rozdielom v postupoch môže byť spôsob aplikácie scenára – niektoré postupy aplikujú do procesného modelu vždy kompletný, v logu zaznamenaný scenár. Iné rozdeľujú scenár na konkrétnejšie udalosti a snažia sa spustením v modeli získať podrobnejší prehľad o vznikajúcich odchýlkach od modelu.

7.1.4. Všeobecnosť

Logy udalostí pochádzajúce z rozličných zdrojov ponúkajú na analýzu často rozsiahle objemy dát, ktoré zachytávajú správanie sa v skúmanej organizácii. Napriek tomu, že tvorba logov prebieha najčastejšie automaticky nie je zaručené, že tieto dátové zdroje budú obsahovať kompletné, systémom umožňované správanie. Takmer s istotou je možné tvrdiť, že udalosti zaznamenané v logoch budú zachytávať len časť možných scenárov. Na uskutočnenie sa niektorých aktivít ešte nemuseli byť vytvorené správne podmienky, systém tieto aktivity ale umožňuje a preto musia byť tieto aktivity zachytené aj v popisnom

procesnom modelu. A práve túto charakteristiku hodnotí kvalitatívne kritérium všeobecnosť.

Modely, ktoré sú ohodnotené vysokou úrovňou presnosti a vhodnosti v porovnaní s logom udalostí strácajú kvalitu práve vo chvíli, kedy je v systéme využitá ďalšia, do vtedy nespustená aktivita. Kritéria presnosti a vhodnosti totiž vyplývajú len zo vzťahu medzi dostupným procesným modelom a zachytenými aktivitami v logu udalostí. Kritérium všeobecnosti predstavuje indikátor toho, či procesný model nie je príliš prispôsobený logu udalostí, ale či popisuje skutočný systém. Hodnota všeobecnosti teda vyjadruje, či a do akej miery je procesný model schopný prezentovať aj správanie sa a akcie, ktoré ešte neboli zachytené v logu a sú umožnené v pozorovanom systéme.

7.2. Hodnotenie algoritmov

7.2.1. Alpha algoritmus

Jedným z najjednoduchších algoritmov, ktoré sú využívané na tvorbu procesných modelov, je alfa algoritmus. Alfa algoritmus je používaný na objavovanie modelu na základe sekvencií krokov z logov udalostí. Objavený model má podobu Petriho siete.

Aplikáciou vyhodnocovacej metodiky je možné určiť kvalitatívne ukazatele algoritmu. Presnosť algoritmu bude v tomto prípade vysoká (> 0.99), čo je spôsobené nízkym počtom možných poradí kombinácií využitých trás. Procesný model teda neumožňuje väčšie množstvo možných scenárov, ktoré by neboli zachytené priamo v logu. Naopak kvalitatívne kritérium všeobecnosti bude dosahovať pomerne nižšej hodnoty (< 0.7). U algoritmu, ktorý pri tvorbe procesného modelu vychádza výhradne z logu udalostí, je ale nižšie hodnoty všeobecnosti možné logicky predpokladať. Keďže je v prezentovanom procesnom modeli nasimulovať takmer každý zaznamenaný scenár, kvalitatívny ukazateľ vhodnosti bude taktiež na vysokej úrovni (> 0.99). Posledným kritériom kvality je jednoduchosť, ktorá je okrem iného ovplyvnená predovšetkým veľkosťou modelu alebo pomerom zduplikovaných aktivít prezentovaných v modeli. V tomto prípade je ale každá aktivita zastúpená v objavenom modeli práve raz, preto model v spojení s algoritmom dosahujú maximálnej úrovne jednoduchosti.

8. Zdroje dát

Výsledky nedávnych výskumov umožňujú vytvoriť predpoveď, ktorá hovorí, že veľkosť digitálneho vesmíru bude každé dva roky minimálne dvojnásobná. Informácie sú uchovávané v databázach, súboroch, správach, error logoch, transakčných logoch, ERP systémoch, dokumentoch, Jednou z hlavných aktivít process miningu je extrakcia práve tých údajov, ktoré majú význam v závislosti na riešenej oblasti a na otázkach, na ktoré sa snažíme získať odpovede.

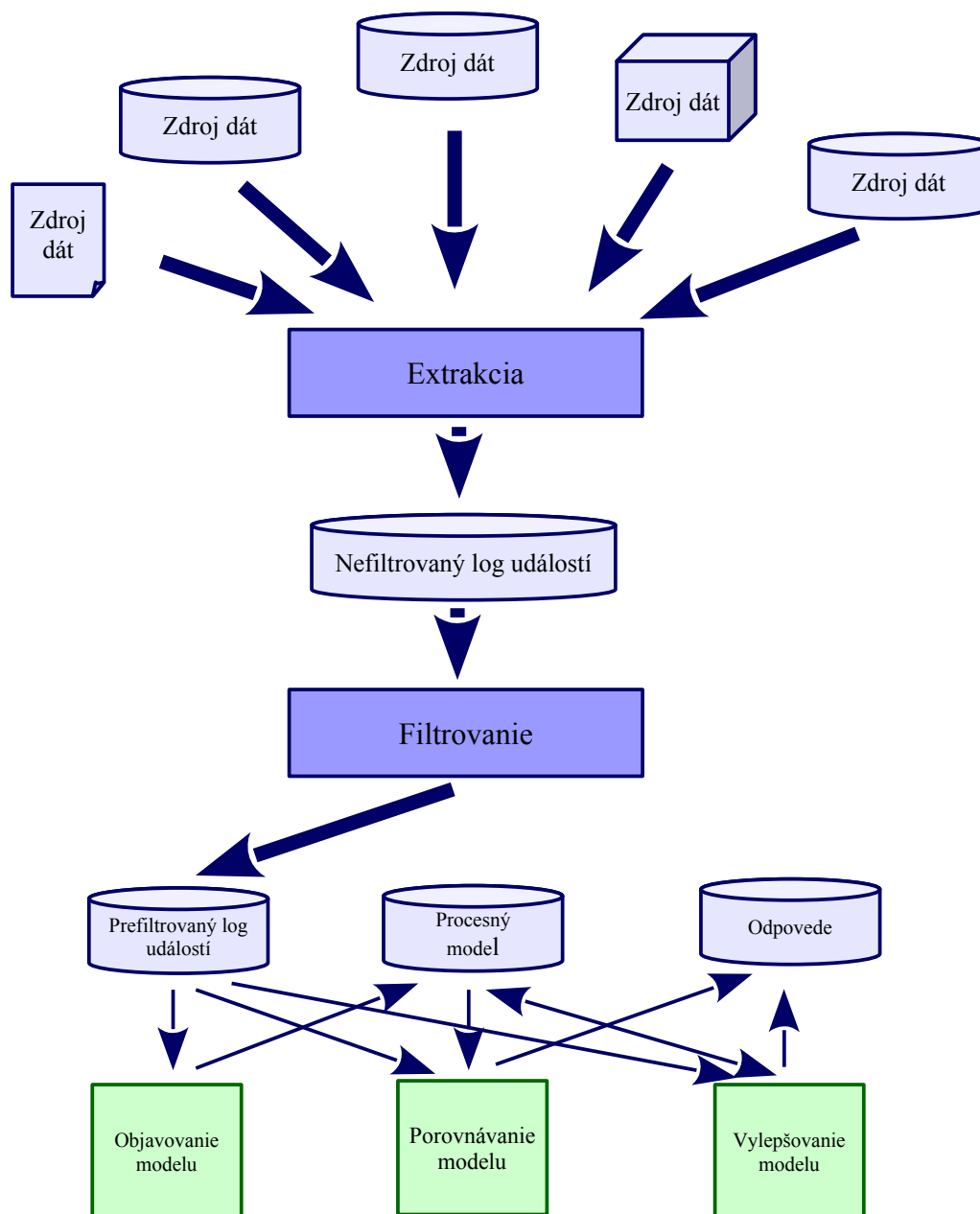
8.1. Analýza zdrojov

Aplikáciou metódik process miningu sa snažíme nájsť odpovede na otázky, ktoré súvisia so skúmanými postupmi a majú teda určitý procesný charakter. Príklad formátu takýchto otázok:

- Aké udalosti nastali v minulosti?
- Aké faktory viedli k vzniku týchto udalostí?
- Aké faktory viedli k odchýlkam oproti plánovanému postupu?
- Je v aktuálne využívanom postupe priestor na vznik problémov?
- Ako je možné upraviť využívaný postup za účelom zvýšenia efektivity?

Analyzované dátové zdroje môžu trpieť rôznymi nedostatkami - nedokonalá štruktúra uchovávaných dát, nejednotné identifikátory objektov, a podobne. Ďalšou prekážkou, na ktorú je možné pri získavaní dát pre process mining naraziť, je veľmi rozsiahly objem dát. Príkladom môže byť ERP systém, ktorý môže obsahovať viac ako 10 000 tabuliek. Extrakcia všetkých týchto dát pravdepodobne ani nebude mať zmysluplný význam pretože je veľmi nepravdepodobné, že na zodpovedanie našich otázok nie je nutné analyzovať taký rozsiahly objem. Informácie vyžadované riešenou úlohou nemusia byť ale súčasťou len jedného zdroja alebo jednej organizácie. V prípade riešenia problematiky s odbavením objednávky bude nutné analyzovať dáta od predajcu ale aj dáta od dopravcu, ktorý zabezpečuje dodanie objednaného tovaru. V ideálnom prípade sú skúmané dáta doplnené o popisné metadata, v mnohých situáciách ale tieto metada nie sú dostupné. Vo výsledku by extrakcia dát potrebných pre analýzu algoritmi process miningu nemala byť riadená možnými zdrojmi a objemom dát ale predovšetkým položenými otázkami [1]. Z čoho vyplýva, že je zbytočné a nevhodné extrahovať a predpripravovať dáta z dostupných zdrojov skôr, ako sú položené príslušné otázky.

... extrakcia dát potrebných pre analýzu algoritmi process miningu by nemala byť riadená možnými zdrojmi a objemom dát ale predovšetkým položenými otázkami ...



Obrázok 8. Schéma znázorňujúce hlavné fázy metodík process miningu [1]

Vo fáze získavania a prípravy potrebných dát je možné využiť postupy aj z prostredia data miningu. Jednou z aplikovateľných best practices z tejto oblasti je tzv. ETL – Extract, Transform, Load[1]. Táto fráza popisuje postupnosť krokov:

1. extrakcie dát z vonkajších zdrojov

2. transformácia dát zo zdrojov do podoby, ktorá je vhodná z hľadiska riešeného problému
3. načítanie dát do cieľového systému, úložiska

Ukladaním vyfiltrovaných informácií z rôznych zdrojov a štruktúr do jediného úložiska dosahujeme postupne stav, v ktorom máme jednoducho k dispozícii pripravené dáta využiteľné pre ďalšiu analýzu a reporty. Zmysluplné zjednotenie dát pochádzajúcich z rôznych systémov a organizácií prináša niekoľko možných prekážok. Rozličné systémy budú napríklad pravdepodobne využívať rozdielne kľúče a jednoznačné identifikátory, informačný webový systém môže používať ako jednoznačný identifikátor užívateľa email. Ekonomický systém ale môže ako identifikátor používať globálny jednoznačný identifikátor GUID. K rozdielom môže dochádzať aj v používanom formátovaní, kedy v jednom zdroji môže byť využívaný formát dátumu v tvare DD-MM-YYYY, a iný zdroj môže mať dátumy uložené vo formáte YYYY/MM/DD. Niektoré organizácie už disponujú určitou formou centrálného úložiska. Je ale nepravdepodobné, že dáta zaznamenané v tomto úložisku poskytnú v prípade potreby nejaké užitočné procesné informácie. Dáta v týchto úložiskách sú uchovávané takým spôsobom, aby umožňovali pohľad na sledovanú doménu z rôznych uhlov a poskytovali vstupy pre vytváranie rôznych reportov.

Bez ohľadu na to, aké zdroje bude nutné spracovať za účelom zodpovedania položených otázok, bude nutné potrebné dáta pretransformovať do logov udalostí. Pri dosahovaní tohto výsledku nie je hlavným problémom samotná syntaktická konverzia do logov ale hlavnou výzvou je určenie zmysluplných a vhodných dát. Jeden dátový zdroj môže poskytovať dáta vhodné pre rôzne riešené oblasti. Z toho vyplýva, že jeden dátový zdroj poskytuje logy udalostí, ktoré zachytávajú rôzne uhly pohľadu na analyzovanú organizáciu. Príkladom je ekonomický systém, ktorý ukladá dáta o zákazníkoch, objednávkach, slúži ako agenda skladových zásob. Tento zdroj poskytuje dáta na sledovanie frekvencie vzniku objednávok za určité obdobia, extrakciou vhodných dát je ale možné využiť zdroj aj na analýzu skladového procesu a analyzovať výsledky spolupráce s dodávateľmi. Riešenie týchto otázok si vyžaduje rozdielne logy udalostí, niektoré informácie budú ale pravdepodobne zaznamenané v oboch logoch.

Spracovanie logu udalostí algoritmami process miningu vedie ale často k vzniku ďalších otázok, ktorých zodpovedanie si vyžaduje spracovanie ďalších dátových zdrojov alebo podrobnejších informácií z už analyzovaných zdrojov, čo má za následok ďalšiu iteráciu extrakcie a filtrovania zmysluplných dát pre log udalostí.

8.2. Proces, prípad, udalosť

Pri aplikovaní metodiky process miningu na vytvorený log udalostí sa predpokladá, že informácie zaznamenané v analyzovanom logu odpovedajú jednému konkrétnemu procesu – dosiahnutie tohto stavu je súčasťou extrakcie a filtrovania dát dostupných z rôznych zdrojov pri vytváraní logu. Súčasne algoritmy miningu analyzujú jednotlivé udalosti z logu s predpokladom, že každá udalosť zachytená v logu patrí jednému konkrétnemu prípadu. Prípad (case) je vlastne jednou inštanciou analyzovaného procesu. Z toho

vyplýva, že jednotlivé záznamy logu musia obsahovať okrem iného identifikátor udalosti, ktorej výskyt vyústil vo vytvorenie záznamu v logu, a identifikátor prípadu, počas ktorého došlo k spusteniu zaznamenanej udalosti. Na základe týchto dvoch atribútov je už možné odsledovať, aké aktivity boli v danom prípade vykonané. Tieto dáta sú vhodné ale skôr na analýzu statickej stránky riešenej domény. S ich pomocou je možné vytvoriť statický pohľad na analyzovanú organizáciu, napríklad v ecommerce oblasti je možné v prípade doplnených identifikátorov konkrétnych entít z týchto logov získať informácie, ktoré produkty sa najčastejšie objednávajú a podobne. Z procesného hľadiska je nutné aby logy obsahovali ešte časový údaj, ktorý poskytuje informáciu o tom, kedy k zalogovanej udalosti došlo. Po doplnení časovej známky sú už k dispozícii všetky potrebné údaje k tomu, aby bolo možné vytvoriť jednotlivé sekvencie udalostí konkrétnych prípadov tak, ako v systéme skutočne nastali.

ID udalosti	Timestamp	Aktivita	ID prípadu
2366784	1425224828	a_incomming	1425224828_A8o6yV7W
2366785	1425224844	a_category	1425224828_A8o6yV7W
2366786	1425224844	a_incomming	1425224844_WGfICAJB
2366787	1425224851	a_search	1425224844_WGfICAJB
2366788	1425224861	a_pl_detail	1425224844_WGfICAJB
2366789	1425224878	a_category	1425224844_WGfICAJB
2366790	1425224907	a_category	1425224828_A8o6yV7W
2366791	1425225053	a_incomming	1425225053_6LecoCfu
2366792	1425225054	a_incomming	1425225054_1jlrzBMw
2366793	1425225061	a_search	1425225054_1jlrzBMw
2366794	1425225065	a_category	1425225053_6LecoCfu
2366795	1425225080	a_pl_detail	1425225054_1jlrzBMw
2366796	1425225104	a_pd_buy	1425225054_1jlrzBMw
2366797	1425225107	a_incomming	1425225107_Xa9sUS6J
2366798	1425225115	a_bf_into_basket	1425225054_1jlrzBMw
2366799	1425225122	a_category	1425225107_Xa9sUS6J
2366800	1425225149	a_category	1425225053_6LecoCfu
2366801	1425225175	a_pl_detail	1425225107_Xa9sUS6J
2366802	1425225186	a_category	1425225107_Xa9sUS6J
2366803	1425225231	a_incomming	1425225231_zB1x2yoZ
2366804	1425225245	a_category	1425225231_zB1x2yoZ
2366805	1425225245	a_category	1425225107_Xa9sUS6J

Obrázok 9. Vzorka logu udalostí v prirodzenej forme

Jednotlivé udalosti zachytené v logu môžu byť doplnené aj o ďalšie doplňujúce atribúty, u udalosti vloženia produktu do košíku to môžu byť doplnkové informácie ako IP zákazníka, ktorý produkt do košíku vložil, identifikátor vloženého produktu, identifikátor kategórie a podobne. Každá udalosť môže mať iné doplňujúce atribúty, vloženie produktu do košíku bude mať iné atribúty ako voľba dopravy v objednávke.

Udalosti a prípady sú reprezentované jednoznačnými unikátnymi identifikátormi, ktoré nám umožňujú označiť konkrétnu udalosť alebo prípad v sekvenciách zachytených v analyzovanom logu

udalostí. Prítomnosť týchto identifikátorov je nevyhnutná pretože mnoho prípadov a aktivít môže mať okrem týchto unikátnych identifikátorov ostatné atribúty rovnaké. Príkladom môžu byť dvaja užívatelia systému, ktorí vykonajú v systéme rovnaké postupnosti krokov s rovnakým výsledkom. V prípade, že tieto dva prípady nie sú jednoznačne odlišené, nie je možné vykonať ich spracovanie z pohľadu dvoch samostatných inštancií analyzovaného procesu. Tieto jednoznačné identifikátory nemusia byť súčasťou informácii z pôvodného dátového zdroja ale môžu byť vygenerované pri extrakcii dát z rôznych zdrojov a sú súčasťou až vytváraného, pre algoritmy process miningu spracovateľného, logu udalostí.

V prípade, že je log udalostí doplnený aj o príslušné role, ktoré zaznamenané udalosti iniciujú, zdroje, ktoré sú spotrebované pri vykonaní udalosti, alebo časový údaj znázorňujúci dobu od poslednej udalosti, je možné s využitím týchto údajov vygenerovať rôzne druhy sekvencií, ktoré budú zachytávať inštanciu daného procesu z rôznych pohľadov.

8.3. Best practices extrakcie a vytvárania logov udalostí

8.3.1. Súvislosť

Pri získavaní informácií z rôznych dátových zdrojov vzniká otázka, ako previazať jednotlivé udalosti z rôznych zdrojov do jednej určitej inštancie procesu. Udalosti v logu udalostí musia byť zoskupené do konkrétneho prípadu. Táto prekážka sa vyskytuje predovšetkým v prípadoch, kedy je nutné analyzovať historické dáta.

8.3.2. Granularita

Dátové zdroje poskytujú informácie s rôznou granularitou vo vzťahu k vykonaným aktivitám [1]. Rôzne pohľady na riešenie doménu môžu vyžadovať prezentovanie reality s rozdielnymi úrovňami podrobností. Pre technického pracovníka organizácie je vhodné prezentovať procesný model výskytu chyby so všetkými podrobnosťami tak, aby model zachytával všetky možné faktory, ktoré mohli viesť k vzniku zalogovanej chyby. Pracovník na základe podrobného procesného modelu navrhuje následne kroky, zmeny konfigurácie systému alebo zmeny v zaužívanom procese za účelom zníženia výskytu faktorov vedúcich k vzniku chyby. Pre vedenie organizácie naopak nie sú dôležité všetky podrobnosti spojené s výskytom nejakej chyby, ktorá je zachytená v error logu. Zmeny a optimalizáciu aktuálne využívaného procesu zmenou počtu zamestnancov na danom oddelení za účelom rýchlejšieho odbavovania vzniknutých problémov vykoná na základe návrhu, ktorý bude obsahovať len počty vzniknutých chýb spolu s prioritou a vážnosťou týchto chýb. Pri extrakcii a príprave logov udalostí z rôznych zdrojov je teda nutné prispôbovať granularitu dostupných informácií práve riešeným otázkam.

8.3.3. Rozsah riešenej oblasti

Rozsiahle systémy poskytujú rozsiahly objem zaznamenaných dát. Dáta môžu byť uložené v desiatkach tisícok tabuliek alebo v stovkách dátových schránok. Vzniká otázka, ktoré dáta je vhodné extrahovať a spracovať do podoby logu udalostí. Odpoveď na túto otázku je možné získať len na základe znalostí problémov z oblasti riešenej organizácie a predovšetkým na základe položených otázok, ktorých odpovede sú získavané aplikáciou process miningu.

8.3.4. Časová známka

Udalosti jednej konkrétnej inštancie skúmaného procesu môžu byť súčasťou rôznych dátových zdrojov. Ak sú udalosti súčasťou jedného zdroja, sú pravdepodobne zaznamenané sekvenčne v takom poradí ako sa v systéme vyskytli a nevzniká tak prekážka pri vytvorení správnej postupnosti udalostí vo výslednom procesnom modeli. V prípade udalostí z viacerých zdrojov je už nutné splnenie podmienky, aby jednotlivé záznamy výskytov udalostí obsahovali aj časovú známku aby bolo možné vytvoriť korektné zoradenú sekvenciu udalostí získaných z odlišných zdrojov.

V tejto situácii ale vzniká niekoľko možných nástrah, napríklad niektoré zdroje skúmaných dát môžu mať oneskorené zaznamenávanie do logu, takže čas uložený u udalosti nie je čas výskytu udalosti ale čas uloženia záznamu o vzniku udalosti do dátového zdroja. V týchto prípadoch, kedy nie je možné spoľahlivo tvrdiť, že všetky časy zaznamenaných udalostí odrážajú skutočne čas výskytu udalostí a nie len čas vzniku záznamu o výskytu udalosti, nie je možné vytvoriť dôveryhodne zoradenú sekvenciu udalostí konkrétnej inštancie procesu. Prostredníctvom inkonzistencie v ukladaní časových známok totiž vzniká priestor na stav, ktorý by mohol vyústiť v procesný model znázorňujúci výskyt následku ešte pred výskytom príčiny. Čiastočným riešením tohto problému je vytváranie modelu za predpokladu len čiastočného zoradovania analyzovaných udalostí.

8.3.5. Podproces

Analyzované logy udalostí pravdepodobne nezachytávajú absolútne celý proces, ktorý prebieha v skúmanej organizácii. Je dôležité si uvedomiť, že informácie zaznamenané v dátovom zdroji zachytávajú pravdepodobne len časť rozsiahlejšieho scenára, ktorý je započatý ešte pred zalogovaním prvej udalosti a bude pravdepodobne pokračovať aj po zalogovaní poslednej udalosti. Vo väčšine prípadov ale máme určenú inicializačnú a ukončovaciu aktivitu, ktoré ohraničujú časť kompletného procesu, ktorá nás zaujíma z pohľadu hľadania odpovedí na položené otázky.

8.4. Štandardné formáty logov udalostí

8.4.1. MXML

V minulosti sa na ukladanie logov udalostí využíval štandard MXML (Mining eXtensible Markup Language). Prostredníctvom tohto štandardu je možné ukladať štruktúrované dáta obsahujúce identifikátory udalostí a prípadov, role, časy, v podobe, ktorá je vhodná pre ďalšie spracovanie

algoritmami process miningu. Ukladanie prebieha s využitím XML syntaxe. MXML poskytuje priestor na ukladanie aj doplnkových informácií udalostí a prípadov. Na základe štandardu MXML vznikol nový formát ukladania logov udalostí – XES [5].

8.4.2. XES

XES vznikol na základe praktických skúseností s využívaním štandardu MXML a od svojho predchodcu sa odlišuje predovšetkým rozsiahlejšími možnosťami rozširovania a je o niečo menej striktný. V roku 2010 bol schválený organizáciou IEEE a stal sa štandardom v oblasti process miningu [5].

Súbor typu XES obsahuje ľubovoľné množstvo zaznamenaných prípadov – inštancií analyzovaného procesu, ktoré sú tvorené postupnosťou udalostí uskutočnených počas vykonávania daného prípadu. Log, jednotlivé prípady, udalosti, všetky z týchto prvkov môžu mať ľubovoľné množstvo zanorených doplnujúcich atribútov. V prípade, že je vhodné použitie nových atribútov s určitou sémantikou, je možné XES doplniť o užívateľské rozšírenie. Pomocou týchto rozšírení je možné popísať sémantiku individuálnych atribútov prvkov zo zpracovávaného logu udalostí. Niektoré atribúty prvkov logu prevedeného do formátu XES sú povinné – prípad musí byť označený jednoznačným unikátnym identifikátorom alebo udalosť konkrétneho prípadu musí mať k dispozícii časový údaj, kedy bola daná udalosť iniciovaná. V súčasnosti je známych niekoľko štandardných rozšírení [5]:

- Rozšírenie obaľujúce atribút *názov*, pre jednoznačnú identifikáciu prípadov a udalostí
- Rozšírenie definujúce *štádium* udalosti z hľadiska životného cyklu udalosti. Možné hodnoty sú *naplánovaná udalosť*, *start udalosti*, *dokončenie udalosti*, *preskočenie udalosti*
- Rozšírenie zachytávajúce *čas*
- Rozšírenie definujúce atribúty udalosti z hľadiska *organizácie*. Tieto atribúty sú *zdroj*, *role*, *skupina*. Ako role môže byť registrovaný zákazník a ako skupina môže byť zákazníci cenovej skupiny VIP

```

<trace>
  <string key="concept:name" value="1425224828_A8o6yV7W"/>
  <event>
    <int key="2366784" value="2366785"/>
    <string key="concept:name" value="a_category"/>
    <string key="lifecycle:transition" value="start"/>
    <date key="time:timestamp" value="2015-03-01T16:47:24.000+01:00"/>
  </event>
  <event>
    <int key="2366784" value="2366790"/>
    <string key="concept:name" value="a_category"/>
    <string key="lifecycle:transition" value="start"/>
    <date key="time:timestamp" value="2015-03-01T16:48:27.000+01:00"/>
  </event>
</trace>
<trace>
  <string key="concept:name" value="1425224844_WGfICAjB"/>
  <event>
    <int key="2366784" value="2366786"/>
    <string key="concept:name" value="a_incomming"/>
    <string key="lifecycle:transition" value="start"/>
    <date key="time:timestamp" value="2015-03-01T16:47:24.000+01:00"/>
  </event>
  <event>
    <int key="2366784" value="2366787"/>
    <string key="concept:name" value="a_search"/>
    <string key="lifecycle:transition" value="start"/>
    <date key="time:timestamp" value="2015-03-01T16:47:31.000+01:00"/>
  </event>
  <event>
    <int key="2366784" value="2366788"/>
    <string key="concept:name" value="a_pl_detail"/>
    <string key="lifecycle:transition" value="start"/>
    <date key="time:timestamp" value="2015-03-01T16:47:41.000+01:00"/>
  </event>
  <event>
    <int key="2366784" value="2366789"/>
    <string key="concept:name" value="a_category"/>
    <string key="lifecycle:transition" value="start"/>
    <date key="time:timestamp" value="2015-03-01T16:47:58.000+01:00"/>
  </event>
</trace>

```

Obrázok 10. Vzorka logu udalostí vo formáte XES

Dostupnosť skutočne vhodných a kvalitných logov udalostí je nevyhnutnou podmienkou na získanie zmysluplných výsledkov po aplikácii metodík process miningu. Vytvorené dôveryhodné logy udalostí môžu byť ale využité aj za iným účelom, napríklad vedenie organizácie môže na základe informácií v logu udalostí sledovať stav a vývoj organizácie v čase a plánovať jej ďalší rozvoj.

9. Real-time process mining

Aplikácia algoritmov process miningu na logy udalostí v offline režime poskytuje podklady pre vytvorenie procesných modelov alebo rozšírenie stávajúcich modelov na základe historických dát. Vďaka technologickému pokroku je v súčasnosti k dispozícii dostatočný vypočetný výkon, ktorý umožňuje spracovanie dát udalostí v reálnom čase, teda v momente kedy táto udalosť nastala. Využitie metodík process miningu v reálnom čase ponúka silný prostriedok na predpovedanie vzniku ďalších udalostí v práve spustenom procese.

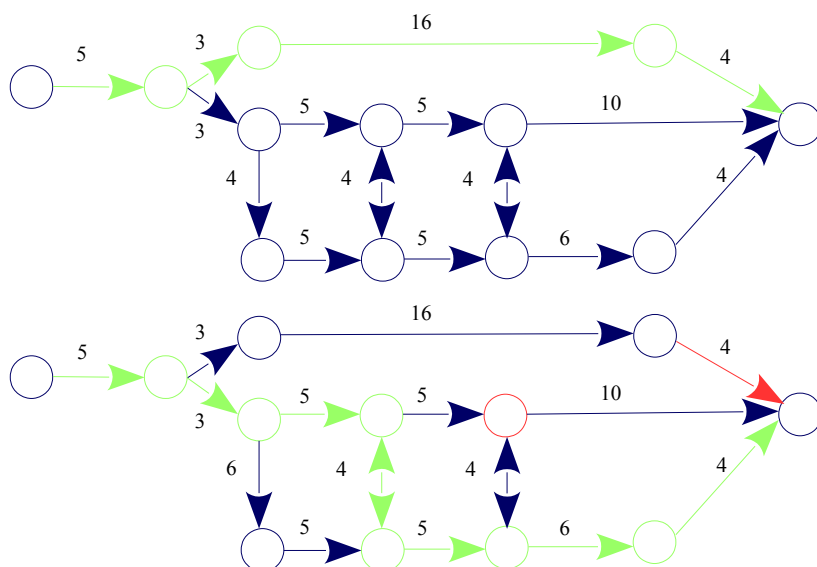
9.1. Pre mortem a post mortem

Na základe definície hlavných oblastí techník process miningu – objavovanie, porovnávanie, vylepšovanie, je možné usúdiť, že analýza vykonávaná algoritmami môže byť iniciovaná online alebo offline. Rozumie sa tým stav, kedy dáta, ktoré spracovávame ako vstupy algoritmov, sú *pre mortem* alebo *post mortem* [1]– scenáre, ktoré pozostávajú zo zaznamenaných udalostí sú už dokončené alebo ešte prebiehajú. *Post mortem* logy udalostí môžu byť využité na analýzu zaužívaného procesu, na získanie podkladov a návrhov na vylepšenie aktuálneho procesného modelu, na vytvorenie zmysluplných

vstupných dát pre audit. Ich využitie ale nie je možné za účelom zmien a vylepšení v aktuálne prebiehajúcej inštancii procesu. Spracovanie *post mortem* logov udalostí je vhodné predovšetkým v offline režime, napríklad za účelom vytvorenia procesného modelu, ktorý bude znázorňovať prebiehajúce inštancie procesu za posledný polrok. Pri spracovaní procesu v online režime je nutné okrem *post mortem* logov udalostí využiť aj aktuálne, *pre mortem* logy udalostí, v ktorých sú zachytené udalosti práve prebiehajúcich scenárov. Historické dáta sú využité predovšetkým na vytvorenie predpovedí vo vykonávaní analyzovaného procesu, v spojení s dátami práve bežiacich inštancií procesu vzniká predikcia zostávajúceho času potrebného na dokončenie vykonávania príslušného prípadu.

9.2. Procesné modely

V oblasti process miningu je možné procesné modely rozlišovať okrem iného podľa zmyslu vytváraného modelu. Rozlišujeme modely, ktoré zachytávajú a prezentujú predpokladaný stav reality - aké kroky, akým spôsobom, akým smerom, by mali byť vykonané. Druhou skupinou sú modely, ktorých hlavnou náplňou je zachytenie stavu reality tak, aká je v skutočnosti.



Obrázok 11. Schéma navigačného procesného modelu. Uzly predstavujú v tomto prípade križovatky cestnej komunikácie. Jednotlivé hrany sú ohodnotené priemerným časom, ktorý si vyžaduje prekonanie tejto cesty. V druhom prípade je červenou farbou znázornená križovatka a cesta, po ktorých nie je možné pokračovať v ceste z dôvodu uzávierky. Navigačný systém vypočítal na základe real-time informácií novú, najefektívnejšiu trasu.

Rozdelením spracovávaných logov udalostí do skupín *pre mortem* a *post mortem*, a analyzovaných modelov do skupín *normatívne* (zachytávajú predpokladaný stav reality) a *popisné* (zachytávajú realitu

tak, aká je v skutočnosti), vznikajú tri skupiny postupov a perspektív process miningu v súvislosti s logmi udalostí a spracovávanými procesnými modelami – kartografia, audit, navigácia [1].

9.2.1. Kartografické procesné modely

„Kartografické“ procesné modely odpovedajú mapám, ktoré popisujú operačný proces v organizácii rovnako ako geografické mapy popisujú určitú geografickú oblasť. Kartografické modely teda nepopisujú predpokladaný stav reality ale skutočný stav reality. Pri tvorbe modelu tohto typu je možné rovnako ako v geografických mapách využívať rôzne hrúbky hrán, napríklad na základe priority väzby alebo frekvencie výskytu javu využitia tejto väzby medzi dvoma aktivitami. Tvorba kartografických procesných modelov pozostáva z troch aktivít: objavovanie, rozširovanie, diagnostika. Aktivita objavovania odpovedá štandardnému postupu vytvárania modelu niektorou z metódik process miningu. Podobne aktivita rozširovania odpovedá aktivite vylepšovania process miningu, počas ktorej dochádza spojením skúmaného logu udalostí a analyzovaného procesného modelu k oprave alebo vylepšeniu existujúceho modelu. Súčasťou tejto aktivity je aj prípadné rozšírenie dostupného modelu o dopĺňujúce atribúty – čas, role, zdroj, ... - vďaka čomu bude takto rozšírený model znázorňovať aj iné perspektívy. Diagnostika kartografického procesného modelu spočíva v analýze prípadných deadlockov alebo v návrhu alternatívnych modelov za účelom získania efektívnejšej varianty analyzovaného modelu.

9.2.2. Audit

Aktivity auditu analyzujú skutočný stav zaužívaných postupov v rámci organizácie za účelom zistenia, či realita odpovedá nastaveným pravidlám, zákonom, požiadavkám manažmentu. Vychádza sa z procesného modelu, ktorý znázorňuje požadovaný stav reality a z logov udalostí, ktoré zachytávajú skutočný stav reality. Uskutočnenie auditu pozostáva z troch fáz: detekcia, kontrola, porovnanie. Vo fáze detekcie sa porovnáva normatívny procesný model a najaktuálnejšie logy udalostí (logy práve prebiehajúcich a pravdepodobne ešte nedokončených inštancií procesu) za účelom odhaľovania odchýlok v reálnom čase. V prípade, že dôjde k odchýlke oproti vopred nadefinovaným pravidlám, je vygenerované upozornenie ešte pred dokončením vykonávania príslušného scenára. Výsledkom fázy kontroly je zistenie kvality analyzovaného procesného modelu, sleduje sa do akej miery je model skutočne dodržaný v spracovávaných logoch udalostí. Tretia fáza aktivít auditu je fáza povyšovania, v ktorej dochádza k povýšeniu pôvodného normatívneho modelu o prvky, ktoré nie sú v pôvodnom modeli prezentované ale z logu udalostí vyplýva, že sú uplatňované v skutočnosti a prinášajú zefektívnenie pôvodne navrhovaného postupu.

9.2.3. Navigačné procesné modely

Navigačné procesné modely prezentujú možné budúcnosti aktuálne prebiehajúcich inštancií procesov a poskytujú užívateľovi priestor na vykonanie práve tých krokov, ktoré budú viesť k požadovanej budúcnosti alebo zmiernia dopady už prebiehajúceho nezastaviteľného nežiadúceho scenára. Navigačným procesným modelom je v skutočnosti aj štandardná navigácia vo vozidle, ktorá na základe historických a aktuálnych dát predpovedá vznik zahltien a doby potrebné na dosiahnutie cieľa pri voľbe rôznych variánt možných trás. Proces tvorby navigačných procesných modelov prechádza tromi fázami,

ktorých výsledkom je dosiahnutie požadovaného cieľa užívateľom procesu. Príklad navigačného modelu je možné vidieť na Obrázku 11. v úvode kapitoly 9.2. Procesné modely. Prvá fáza je fáza prieskumu, v ktorej analyzujeme aktuálne dáta z logov udalostí a porovnávame ich s historickými dátami, s inštanciami procesu, ktoré prebehli v minulosti. Vo fáze predikcie využívame výsledky z predchádzajúcej fázy na tvorbu predpovedí o možnej budúcnosti, napríklad nutný zostávajúci čas na dokončenie aktuálne bežiackej inštancie analyzovaného procesu. Fáza doporučení pozostáva z návrhov práve tých krokov, ktoré vedú k najefektívnejšiemu a najšetrnejšiemu dosiahnutiu posledného kroku práve bežiaceho scenára.

9.3. Operačný support

Tradičný pohľad na vytváranie a analýzu procesných modelov spočíva vo využití historických dát, z čoho vyplýva, že analyzované logy udalostí zachytávajú už ukončené inštancie procesu. Pri tvorbe a rozširovaní procesných modelov je ale možné využiť aj dáta aktuálne bežiacich inštancií procesu a poskytovať odpovede na položené otázky priebežne. Operačný support sa zameriava práve len na analýzu tých scenárov, ktoré ešte nie sú dokončené a teda existuje možnosť ovplyvniť ich pokračovanie. Operačný support pozostáva z troch kľúčových krokov – detekcia, predikcia, doporučenie.

9.3.1 Detekcia

V tomto kroku dochádza k porovnaniu aktuálne už prebehnutej časti analyzovanej inštancie procesu so stabilným procesným modelom. Výsledkom porovnania môže byť zistenie výskytu udalosti na mieste, na ktorom nebola predpokladaná alebo nie je povolená vzorovým modelom. V tomto prípade bude vygenerované upozornenie. Príkladom môže byť detekcia a zasielanie chýb riešiteľovi v závislosti na príčine chyby. Po zaznamenaní udalosti, ktorá predstavuje vznik chyby v systéme je nutné preveriť možný vznik príčiny na serveri A, možný vznik príčiny na serveri B, možný vznik príčiny na základe užívateľskej akcie, a až následne sa riešenie chyby odosiela riešiteľovi, ktorý vyplýva z príčiny zaznamenatej chyby a sú vykonané ďalšie aktivity vyplývajúce z predpísaného postupu. V prípade, že systém operačného supportu obdrží informáciu o výskytu udalosti odoslania riešiteľovi ešte pred tým, ako zaznamená výskytu udalosti kontroly serveru A, kontroly serveru B, kontroly užívateľskej akcie, bude systémom detekované odchyľka oproti normatívnemu procesnému modelu.

9.3.2 Predikcia

Na základe upozornení z kroku detekcie je v ďalšom kroku vykonaná predpoveď zostávajúceho času nutného pre riadne dokončenie scenára alebo predpoveď zdrojov, ktoré si dokončenie scenára bude vyžadovať. Predpoveď je založená predovšetkým na historických dátach, na logoch udalostí, ktoré zachytili podobné, nedokončené inštancie analyzovaného procesu v minulosti. ďalšími príkladmi možných predpovedí môže byť aj pravdepodobnosť úspešného dokončenia aktuálne bežiackej inštancie procesu, pravdepodobnosť výskytu určitej aktivity, pravdepodobnosť nutnosti zapojenia konkrétnej osoby do aktuálne prebiehajúceho scenára.

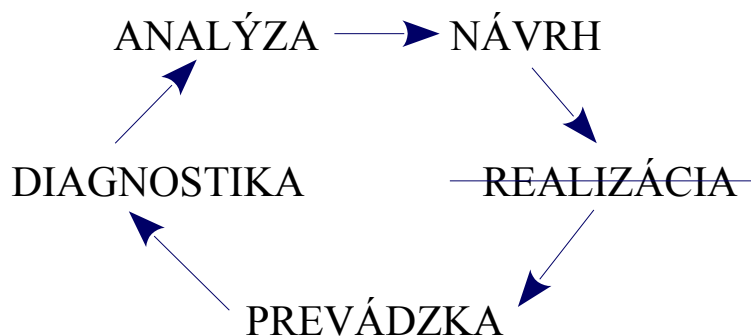
9.3.3 Doporučenie

Na základe predpovedí a historických dát z logov udalostí sú vykonané odporúčenia, aké ďalšie aktivity má vykonať príslušná entita. Operačný support má pri tvorbe odporúčení k dispozícii rozsah možných rozhodnutí práve skúmanej entity a procesný model vytvorený z historických logov udalostí. Pri analýze odporúčení vzniká takzvaný model odporúčení, ktorý znázorňuje ďalšie možné kroky aktuálne bežiackej inštancie normatívneho procesného modelu. Model odporúčení môže byť odlišný pre každú inštanciu v závislosti na výskyte odchýlok pri vykonávaní jednotlivých scenárov. Doporučenie vydané operačným supportom by malo viesť k najefektívnejšiemu dosiahnutiu požadovaných cieľov a k vyriešeniu riešených úloh.

Niekoľko odporúčení na nasledujúci krok sa môže odlišovať numerickým ukazateľom, ktorý indikuje pravdepodobnosť, že po zvolení práve tejto varianty ďalšieho kroku bude cieľ dosiahnutý v požadovanej kvalite. Príkladom je navigácia vo vozidle, ktorá môže indikovať možné smery zmeny jazdy spolu s numerickým ukazateľom, ktorý vyjadruje aká je pravdepodobnosť, že sa po zvolení daného smeru dostane vozidlo do cieľového bodu v nastavenom čase. Dosiahnutie cieľa alebo splnenie riešenej úlohy môže byť závislé aj na ďalších udalostiach, ktoré môžu v organizácii (v doprave) nastať a preto vo väčšine prípadov nie je možné poskytnúť odporúčenie s pravdepodobnosťou 100% dosiahnutia cieľa v požadovanej kvalite (v prípade navigácie je požadovanou kvalitou dosiahnutie cieľa v nastavenom čase). V organizácii je možné sledovať najčastejšie odporúčenia, ktoré sú postavené na základe cieľov typu: Zníženia objemu spotrebovaných zdrojov, minimalizácia škody po výskyte odchýlky, zníženie času potrebného na splnenie určitej úlohy.

10. Vymedzenie uplatnenia proces miningu v biznis procese

V predchádzajúcich kapitolách sme dokázali, že práca nad dátami v process miningu nie je uplatniteľná len v offline režime. V skutočnosti je ale využitie možné v takmer každej fáze biznis procesu – okrem fázy implementácie [1].



Obrázok 12. Fázy biznis procesu, v ktorých je možné a vhodné uplatniť metodiky process miningu

Používaním miningu priamo v online režime získavame výhody kontinuálnej kontroly, ktorá nám umožňuje odhaľovanie odchýlok a problémov práve vo chvíli, keď k nim dochádza a je možné teda reagovať okamžite, čo môže mať za následok výrazné zníženie prípadných škôd. Kontinuálna kontrola nám za pomoci dát získaných v minulosti taktiež umožňuje priebežné určovanie zostávajúceho času práve bežiacich úloh.

- Fáza analýzy – s využitím techník process miningu môžeme skúmať analýzou vytváraný model a zároveň vytvárať alternatívne modely.
- Fáza návrhu – v tejto fáze sa môžeme zamerať hlavne na rozširovanie a vylepšovanie navrhnutého modelu.
- Fáza provozu – uplatnenie process miningu môže zabezpečiť monitorovanie priebehu a získavanie potrebných dát pre ďalší support. Súčasne je možné tieto dáta využiť k menším prispôbeniam.
- Fáza diagnostiky – dáta z miningu sú v tejto fáze využívané predovšetkým k rozsiahlejším úpravám v procese, výsledky tejto aplikácie môžu znovu vstupovať napríklad do fázy návrhu.

11. Metodika a verifikácia procesných modelov

11.1. Problém

V súčasnosti existuje niekoľko techník a postupov na verifikáciu a analýzu procesných modelov. Tieto techniky sú založené na hľadaní odpovedí na konkrétne otázky, ktorých zodpovedanie poskytne vstup na následné rozšírenie a zdokonalenie pôvodného modelu.

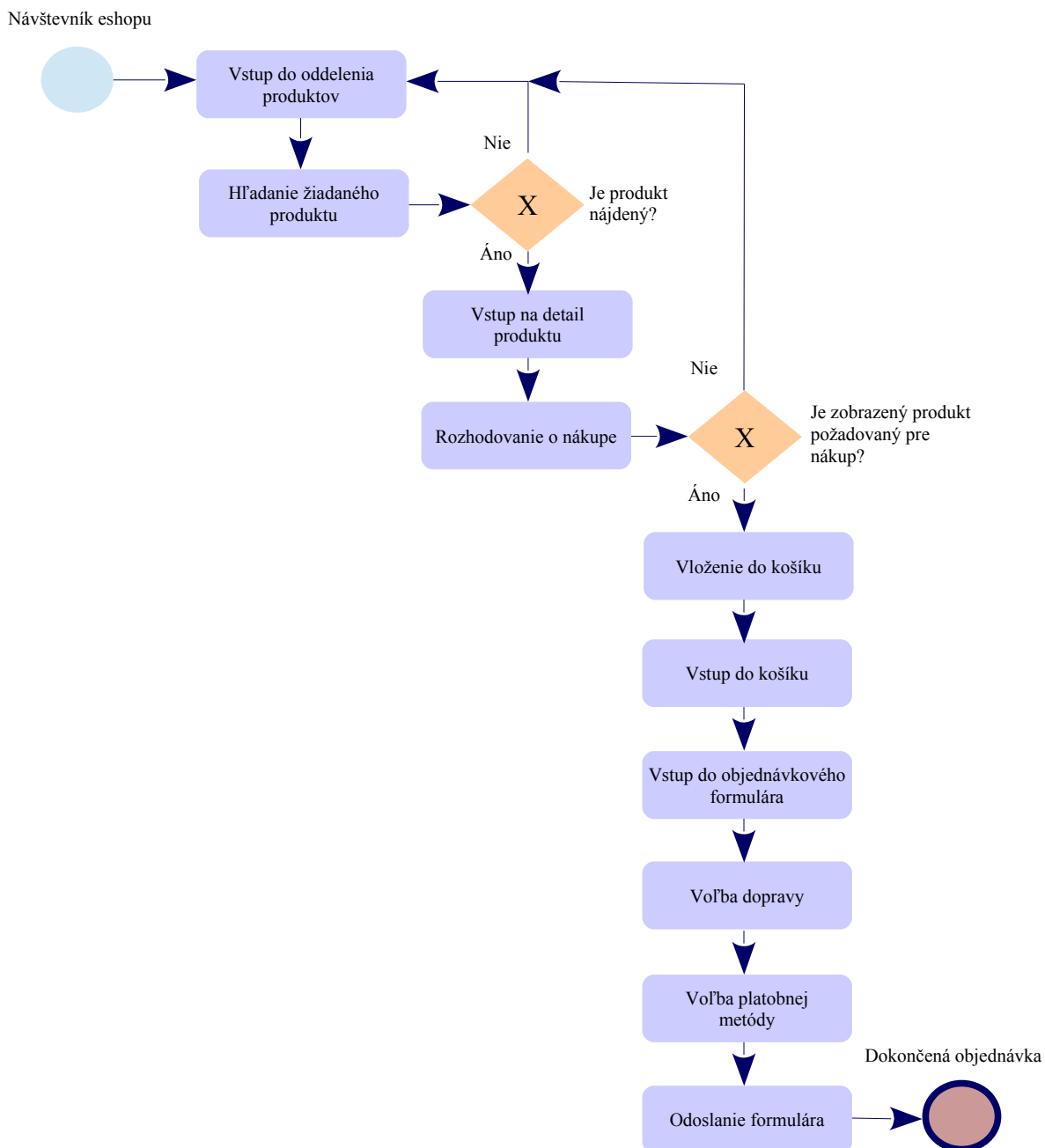
Verifikácia procesných modelov predstavuje proces zodpovedania hlavnej otázky “Odpovedá dostupný model procesu realite?”. Pôvodný model môže byť vytvorený manuálne ešte pred implementáciou na základe požiadavkov stakeholderov a na základe plánovaného stavu reality. Takto vytvorený model môže poskytovať informácie potrebné pre hrubý prehľad procesov v organizácii alebo v systéme, jeho použitie bude ale pravdepodobne zdrojom problémov v prípade auditu alebo plánovania rozširovania organizácie o ďalšie aktivity. Využívanie neaktuálneho normatívneho procesného modelu môže mať za následok aj výrazne mrhanie zdrojov – napríklad v organizácii zaoberajúcej sa logistikou dôjde k implementácii rozšírenia funkcionality automatizovaného prenosu čísla balíka zásielky do fakturačného dokladu, pritom v skutočnosti táto funkcionality nie je vôbec z určitého dôvodu využívaná (nepohodlie / neistota / nedokonalosť) a užívateľ vkladá číslo zásielky do dokladu ručne. V tomto prípade došlo k implementácii funkcionality, ktorú pravdepodobne nebude takmer nikto využívať.

Zastaralý procesný model nemôže byť ani využitý na prezentáciu osvedčených a efektívnych postupov aplikovaných v danej organizácii. V tomto pôvodnom modeli totiž tieto reálne aplikované a až

postupne časom zdokonalené postupy a metodiky pravdepodobne nie sú zachytené. V tejto chvíli je nutné existujúci procesný model nejakým spôsobom verifikovať a navrhnúť zmeny, ktoré budú viesť k zdokonaleniu zastaralého modelu alebo zmeny, ktoré budú viesť k zmene v analyzovanej organizácii alebo systéme. V niektorých prípadoch sú totiž odchýlky využívané v skúmanej organizácii nie výsledkom efektívneho a osvedčeného postupu ale skôr neznalosťou užívateľa alebo nedokonalosťou používaného systému a tieto neefektívne a nežiadúce „skratky“ môžu viesť k problémom a nestabilite v budúcnosti organizácie.

*... je nutné existujúci procesný model nejakým spôsobom verifikovať a navrhnúť zmeny,
ktoré budú viesť k zdokonaleniu zastaralého modelu alebo zmeny, ktoré budú viesť
k zmene v analyzovanej organizácii alebo systéme ...*

V našom prípade je riešenou organizáciou prostredie ecommerce a konkrétnym analyzovaným procesom, na ktorom overíme výsledky vytvorenej metodiky verifikácie procesu, je proces objednávky. Na implementáciu navrhovaného riešenia využijeme aplikáciu ProM.



Obrázok 13. Schéma procesného modelu - Objednávkový proces od vstupu do systému po odoslanie objednávkového formulára

11.2. Vstupné dáta

Na umožnenie verifikácie procesného modelu voči stavu reality budeme potrebovať:

- Vizualnú prezentáciu pôvodného procesného modelu

- Otázky na ktoré sa verifikáciou snažíme nájsť odpovede
- Dáta, ktoré zachytávajú skutočný obraz reality v analyzovanej organizácii

11.2.1. Pôvodný procesný model

Vlastník ecommerce riešenia poskytol na analýzu pôvodné schéma procesného modelu (Obrázok 13), ktoré zachytáva predpokladaný scenár návštevy užívateľa a prechod objednávkovým procesom od vstupu do eshopu až po odoslanie formulára objednávky.

Procesný model zachytáva len základný scenár objednávkového procesu. Je pravdepodobné, že možná postupnosť aktivít v systéme nebude odpovedať vždy presne postupnosti prezentovanej uvedeným modelom – napríklad je možné opustiť objednávkový formulár a prejsť znovu do kategórie produktov. Pre návrat do košíku nie je nutný nákup ďalšieho produktu. A podobne.

11.2.2. Otázky, na ktoré sa snažíme nájsť odpovede verifikáciou procesného modelu

Prostredníctvom konzultácii a dotazovania sa s vlastníkom ecommerce riešenia spolu s analýzou schématu pôvodného procesného modelu vzniklo niekoľko možných dotazov, na ktoré odpovede sa verifikáciou modelu pokúsime nájsť odpovede. Otázky:

1. Odpovedá dostupný procesný model skutočnému stavu reality?
2. Je pravda, že objednávku je možné dokončiť až po vložení nejakého produktu do košíku?
3. Je zvolený druh platobnej metódy závislý na zvolenej doprave?
4. Je po dokončení objednávky zákazníkovi odoslaný email s potvrdením objednávky?
5. Je možné dokončiť objednávku s tovarom v celkovej cene 0 korún?
6. Je možné vykonať objednávku bez návštevy kategórie produktov?
7. Je pravda, že každá dokončená objednávka obsahuje dopravu?
8. Je umožnené zákazníkovi uplatniť po príchode do objednávkového formulára zľavový kupón?

Pri dôkladnejšom pozorovaní položených otázok je jasné, že niektoré z otázok nemajú náväznosť na analyzovaný procesný model. Tieto otázky popisujú aktivity, ktoré nemajú náväznosť na aktivity zachytené v dostupnom modeli a preto nie je možné získať odpovede na tieto otázky výhradne prostredníctvom verifikácie existujúceho modelu. Otázky, na ktoré nie je možné získať odpovede verifikáciou dostupného modelu:

4. Je po dokončení objednávky zákazníkovi odoslaný email s potvrdením objednávky?
- táto otázka síce má procesný charakter, aktivity, ktoré popisuje majú ale presah za aktuálne modelom prezentovaný scenár aktivít objednávkového procesu. Odpoveď na túto otázku by bolo možné získať rozšírením existujúceho modelu o aktivitu odoslania emailu a verifikáciou nového modelu
5. Je možné dokončiť objednávku s tovarom v celkovej cene 0 korún?
- dostupný procesný model neobsahuje doplňujúce atribúty ako cena vkladaneho produktu a podobne. Odpoveď na túto je možné ale získať priamo, bez nutnej verifikácie modelu a to vyhľadáním objednávky s nulovou cenou priamo v agende objednávok. Zároveň samotná otázka nemá procesný charakter – nedotazuje sa na nejakú predpokladanú postupnosť aktivít nákupného procesu.
8. Je umožnené zákazníkovi uplatniť po príchode do objednávkového formulára zľavový kupón?
- táto otázka síce má procesný charakter, aktivity, ktoré popisuje ale nemajú možnú väzbu na modelom prezentovaný scenár aktivít objednávkového procesu. Odpoveď na túto otázku by bolo možné získať rozšírením existujúceho modelu o aktivitu uplatnenia zľavového kupónu a verifikáciou nového modelu

11.2.3. Dáta, ktoré zachytávajú skutočný obraz reality v analyzovanej organizácii

Ako vstupné dáta nám poslúžia dáta z heatmapy, ktorá bola na riešení implementovaná v minulosti za účelom marketingovej analýzy. Heatmapa zachytávala priame užívateľské aktivity kliknutí na prvky systému. Súčasťou dát sú aj niektoré aktivity objednávkového procesu.

Voľba zdroja dát sa riadi z poznatkami a best practices z oblasti process miningu:

- Voľba zdroja dát výsledkom analýzy položených otázok na ktoré sa snažíme nájsť verifikáciou odpovede
- Voľba zdroja dát ideálne reflektuje riešenie organizáciu, heatmapy obsahujú čisto len aktivity spojené s ecommerce riešením a majú uložené informácie majú procesný charakter
- Dáta sú z jedného zdroja, odpadajú tak nástrahy spojené s nejednoznačnými identifikátormi
- Zvolený zdroj dát obsahuje všetky potrebné informácie, ktoré sú vyžadované za účelom vytvorenia konkrétnych inštancií analyzovaného procesu (prípady)
- Jednotlivé záznamy obsahujú časové údaje, nie je nutné sa teda spoliehať len na sekvenčné ukladanie vyskytnutých udalostí

- Dáta z heatmapy je možné ohodnotiť štvrtým stupňom kvality dát (piaty stupeň označuje najkvalitnejšie dáta) – dáta sú zaznamenávané automaticky, systematicky, dôveryhodne, a je možná extrakcia úplných scenárov
- S využitím udalostí zachytených prostredníctvom heatmapy je možné vizualizovať skutočný stav reality

ID udalosti	Timestamp	Aktivita	ID prípadu
2366784	1425224828	a_incomming	1425224828_A8o6yV7W
2366785	1425224844	a_category	1425224828_A8o6yV7W
2366786	1425224844	a_incomming	1425224844_WGfICAjB
2366787	1425224851	a_search	1425224844_WGfICAjB
2366788	1425224861	a_pl_detail	1425224844_WGfICAjB
2366789	1425224878	a_category	1425224844_WGfICAjB
2366790	1425224907	a_category	1425224828_A8o6yV7W
2366791	1425225053	a_incomming	1425225053_6LecoCfu
2366792	1425225054	a_incomming	1425225054_1jlrzBMw
2366793	1425225061	a_search	1425225054_1jlrzBMw
2366794	1425225065	a_category	1425225053_6LecoCfu
2366795	1425225080	a_pl_detail	1425225054_1jlrzBMw
2366796	1425225104	a_pd_buy	1425225054_1jlrzBMw
2366797	1425225107	a_incomming	1425225107_Xa9sUS6J
2366798	1425225115	a_bf_into_basket	1425225054_1jlrzBMw
2366799	1425225122	a_category	1425225107_Xa9sUS6J
2366800	1425225149	a_category	1425225053_6LecoCfu
2366801	1425225175	a_pl_detail	1425225107_Xa9sUS6J
2366802	1425225186	a_category	1425225107_Xa9sUS6J
2366803	1425225231	a_incomming	1425225231_zB1x2yoZ
2366804	1425225245	a_category	1425225231_zB1x2yoZ
2366805	1425225245	a_category	1425225107_Xa9sUS6J

Obrázok 14. Události zaznamenané prostredníctvom heatmapy. ID prípadu je vlastne jednoznačný identifikátor návštevníka.

Pred použitím udalostí z heatmapy v aplikácii ProM bude nutné časový údaj prekonvertovať z timestampu do formátu mm.dd.YYYY HH:ii:ss.

11.3. Analýza a návrh

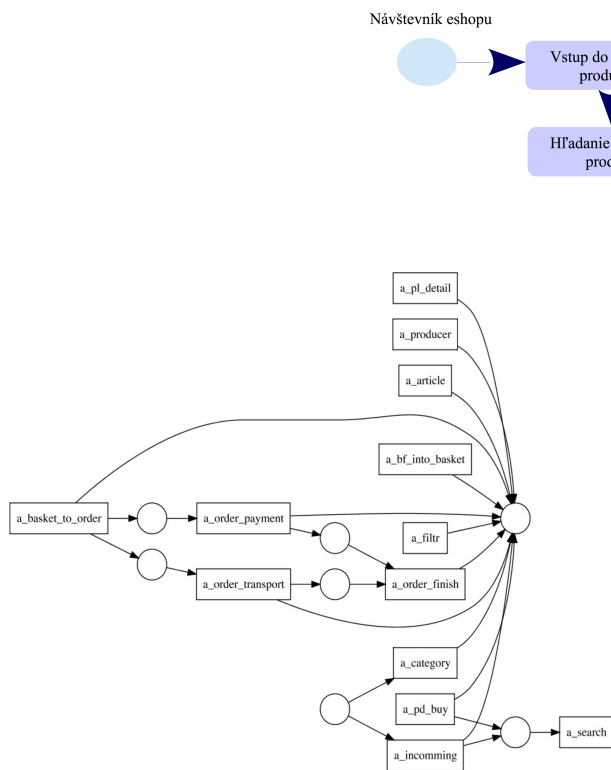
11.3.1. Skúmanie a experimenty existujúcich možností analýzy modelu

Hlavnou náplňou tejto práce je okrem vytvorenia novej metodiky na analýzu procesných modelov aj pochopenie existujúcich postupov a algoritmov process miningu. V pyramíde úrovní osvojenia si schopností process miningu sú na prvých miestach fázy porozumenia a aplikácie poznatkov process

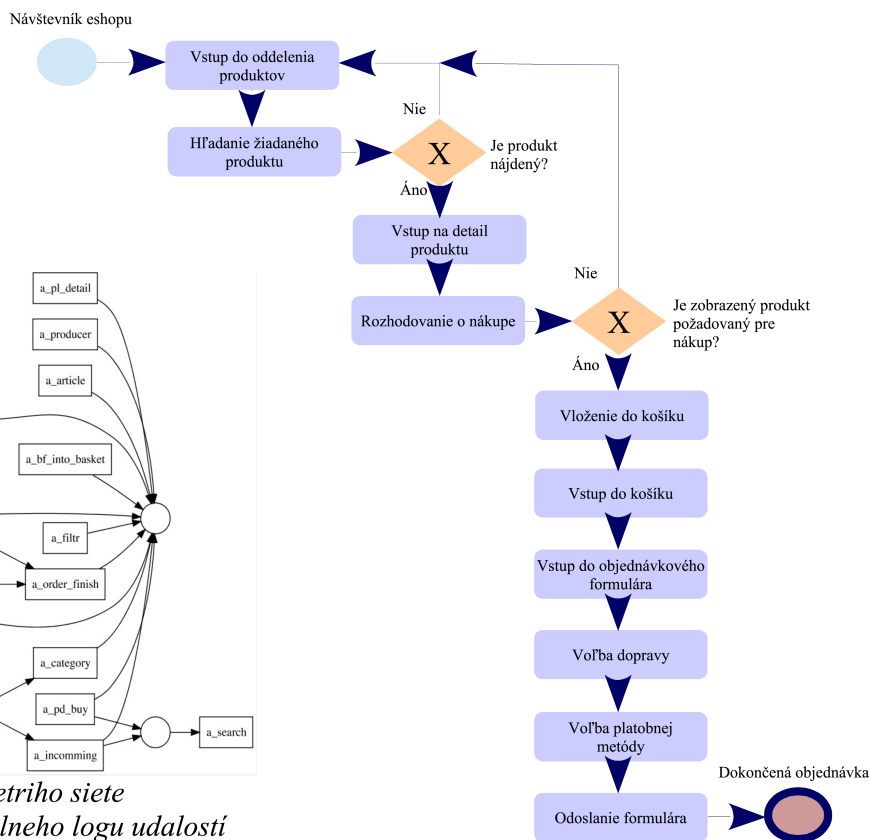
miningu. V tejto kapitole si teda overíme niektoré existujúce algoritmy, ktoré sú implementované v aplikácii ProM.

- **Alpha Miner**

- Plugin implementuje kolekciu algoritmov, z rodiny alpha algoritmov. Na vstupe preberá log udalostí a výsledkom je určitá forma Petriho siete. Túto Petriho sieť môžeme porovnať s našim pôvodným procesným modelom



Obrázok 16. Schéma formy Petriho siete vygenerovanej na základe reálneho logu udalostí analyzovaného procesu

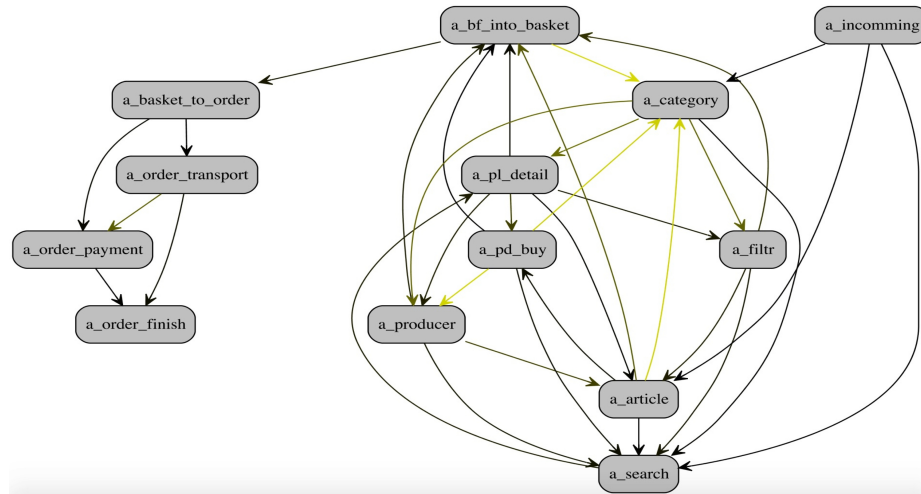


Obrázok 15. Schéma pôvodného modelu procesu

- Z modelu automaticky vygenerovaného Alpha minerom je možné vyčítať niektoré podprocesy, napríklad proces od príchodu do objednávkového formulára až po dokončenie objednávky, alebo fakt, že niektorí užívatelia opustili systém po príchode do objednávkového formulára a ešte pred dokončením objednávky. Z vytvoreného schématu ale nie je možné získať odpovede na položené otázky, napríklad nie je možné odpovedať na otázku, či je možné vytvoriť objednávku bez toho aby došlo najprv k vloženiu produktu do košíka.
- Alpha algoritmus nie je vhodný pre použitie v reálnej situácii z dôvodu neschopnosti rozumného vysporiadania sa s neúplnými scenármi zachytenými v logu. Využíva sa predovšetkým na teoretické znázornenie algoritmického vytvárania procesného modelu.

- **Objavovanie grafu**

- Plugin, ktorý umožňuje vytvorenie jednoduchého grafu aktivít vyplývajúcich z analyzovaného logu udalostí.



Obrázok 17. Triviálny graf aktivít na základe reálneho logu udalostí analyzovaného procesu

- Pozorovaním grafu je možné vysloviť tvrdenie o odpovedi na otázku „Je zvolený druh platobnej metódy závislý na zvolenej doprave?“. Odpoveď, ktorú poskytuje jednoduchý graf je „Nie“. V niektorých prípadoch došlo očividne najprv k voľbe platobnej metódy a až následne k voľbe dopravy. Na poradí teda nezáleží a nie je teda možné tvrdiť, že platobná metóda je závislá na zvolenej doprave. Získavanie odpovedí priamym pozorovaním môže viesť ale k nekorektným výsledkom, ktoré sú spôsobené nesprávnym pochopením grafu hlavne u komplexnejších analyzovaných procesov.

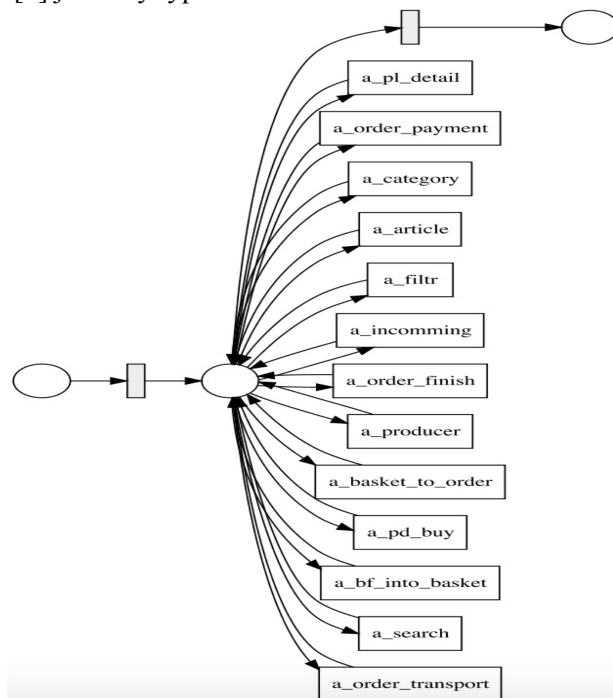
- **Matica vzťahov aktivít**

- Plugin vytvára na základe logu udalostí maticu vzťahov medzi jednotlivými aktivitami. Hodnota 1 značí, že je možné potvrdiť jasný vzťah medzi týmito dvoma aktivitami. Hodnota -1 značí, že z logov vyplýva nulový vzťah medzi dvoma aktivitami. Hodnota 0 značí, že na základe dostupných údajov nie je možné určiť, či vzťah existuje alebo neexistuje.

Matrix	a_article	a_basket_to_order	a_bf_into_basket	a_category	a_filtr	a_incoming	a_order_finish	a_order_payment	a_order_transport	a_pd_buy	a_pl_detail	a_producer	a_search
a_article	0.0	-1.0	0.6666666666666666	0.18674698...	-1.0	-1.0	-1.0	-1.0	-1.0	0.94117647...	-1.0	-1.0	0.98795186...
a_basket_to_order	-1.0	-1.0	-0.9051094...	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	0.99135299...	-1.0	-1.0	-1.0
a_bf_into_basket	-1.0	0.9051094...	0.0	0.15730337...	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
a_category	-0.18674698...	-1.0	-0.15730337...	0.0	0.62280701...	-1.0	-1.0	-1.0	-1.0	-0.1935483...	0.57789855...	0.60526315...	0.98412698...
a_filtr	0.875	-1.0	0.8	-0.62280701...	-1.0	-1.0	-1.0	-1.0	-1.0	0.0	-1.0	-1.0	0.90909090...
a_incoming	0.99135299...	-1.0	-1.0	0.99874529...	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	0.99833610...
a_order_finish	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
a_order_payment	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
a_order_transport	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
a_pd_buy	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
a_pl_detail	0.94117647...	-1.0	0.99135299...	0.99874529...	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	0.98795186...
a_producer	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
a_search	0.98795186...	-1.0	-1.0	0.98412698...	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0

Obrázok 18. Matica vzťahov aktivít

- Z matice vzťahov vyplýva, že je možné potvrdiť veľmi výraznú väzbu medzi aktivitami príchodu do objednávkového formulára a aktivitou voľby dopravy a voľby platobnej metódy. Súčasne je veľmi silná väzba medzi aktivitami voľba platobnej metódy a dopravy a aktivitou dokončenia objednávkového procesu. Na aktivitu vloženia produktu do košíku majú silné väzby aktivity filtrovania podľa výrobcu, zobrazenie detailu produktu, zobrazenie oddelenia produktov, ale aj zobrazenie článku (môžeme predpokladať, že súčasťou článku sú aj ponúkané akčné produkty, na ktoré má daný článok upozorniť a motivovať ľudí k nákupu)
- **Vytvorenie flower modelu (kvetinový model)**
 - Flower model [4] je určitý typ Petriho siete.



Obrázok 19. Flower model vygenerovaný na základe analyzovaného logu udalostí

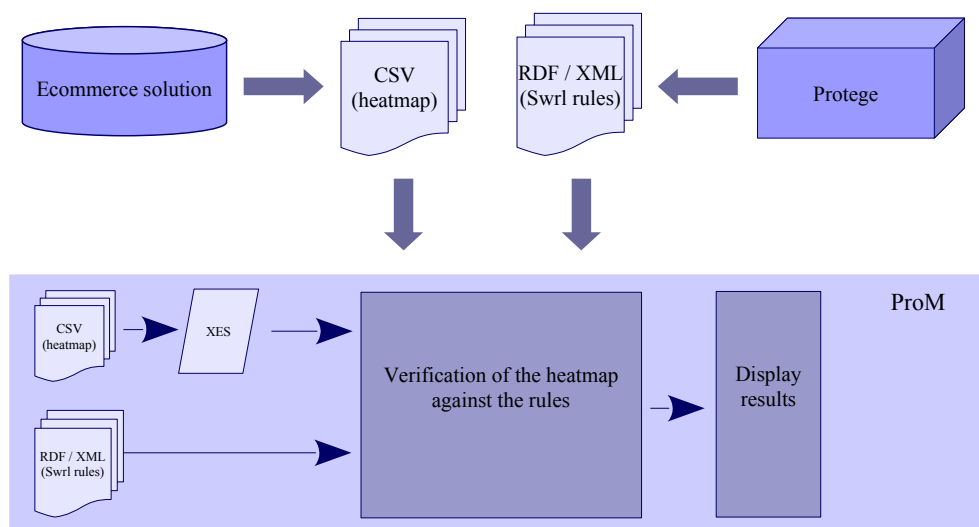
- Pri vizuálnej analýze je možné určiť pozitívne stránky tohto typu procesného modelu, ktorými sú vysoká úroveň jednoduchosti a vhodnosti. Model je jednoduchý na vizuálnu analýzu a určite povoľuje simuláciu všetkých, v logu zaznamenaných, scenárov (teda aj scenáre obsahujúce šum). Už na prvý pohľad sú ale jasné aj nedostatky tohto typu modelu, a to veľmi nízka presnosť pretože model umožňuje veľké množstvo scenárov, ktoré nie sú zachytené v logu udalostí a nie sú umožnené organizáciou.
- **BPMN analýza s využitím Casual Net Mineru**
 - Tento plugin slúži na vytvorenie grafického znázornenia procesov zachytených v logu udalostí podľa normy BPMN.

- $A_order_finish(?a_order_finish) \wedge isOrderFinishedByCreatedBasket(?a_order_finish, ?a_basket_to_order) \wedge isCreatedBasketWithBoughtProduct(?a_basket_to_order, ?a_pd_buy) \rightarrow isOrderFinishedWithBoughtProduct(?a_order_finish, ?a_pd_buy)$
- $A_order_finish(?a_order_finish) \wedge isOrderWithPaymentMethod(?a_order_finish, ?a_order_payment) \wedge isPaymentMethodDependentOnTransport(?a_order_payment, ?a_order_transport) \rightarrow isOrderWithTransport(?a_order_finish, ?a_order_transport)$
- $A_order_finish(?a_order_finish) \wedge isOrderWithPurchasedProduct(?a_order_finish, ?a_pd_buy) \wedge isPurchasedProductOnProductDetail(?a_pd_buy, ?a_pl_detail) \wedge isDisplayedProductDetailFromVisitedCategory(?a_pl_detail, ?a_category) \rightarrow isCreatedOrderAfterVisitedCategory(?a_order_finish, ?a_category)$

V rámci implementácie novej metodiky verifikácie procesu bude nutné implementovať v nástroji ProM aj parser na parsrovanie pravidiel vyexportovaných z nástroja Protege. Aktuálne nástroj ProM nedisponuje funkcionalitami na prácu s pravidlami.

11.3.3. Popis postupu navrhovanej metodiky

Na uskutočnenie verifikácie procesného modelu máme k dispozícii logy udalostí, ktoré zachytávajú skutočný stav reality a formálny zápis otázok, ktoré vyplývajú z analyzovaného procesného modelu, a na ktoré sa verifikáciou pokúsime nájsť odpovede. Postup spracovania analyzovaných dát a proces verifikácie prípadov zachytených v logu udalostí sú vizuálne znázornené na Obrázku 21 a na Obrázku 23.



Obrázok 21. Schéma postupu spracovania analyzovaných dát

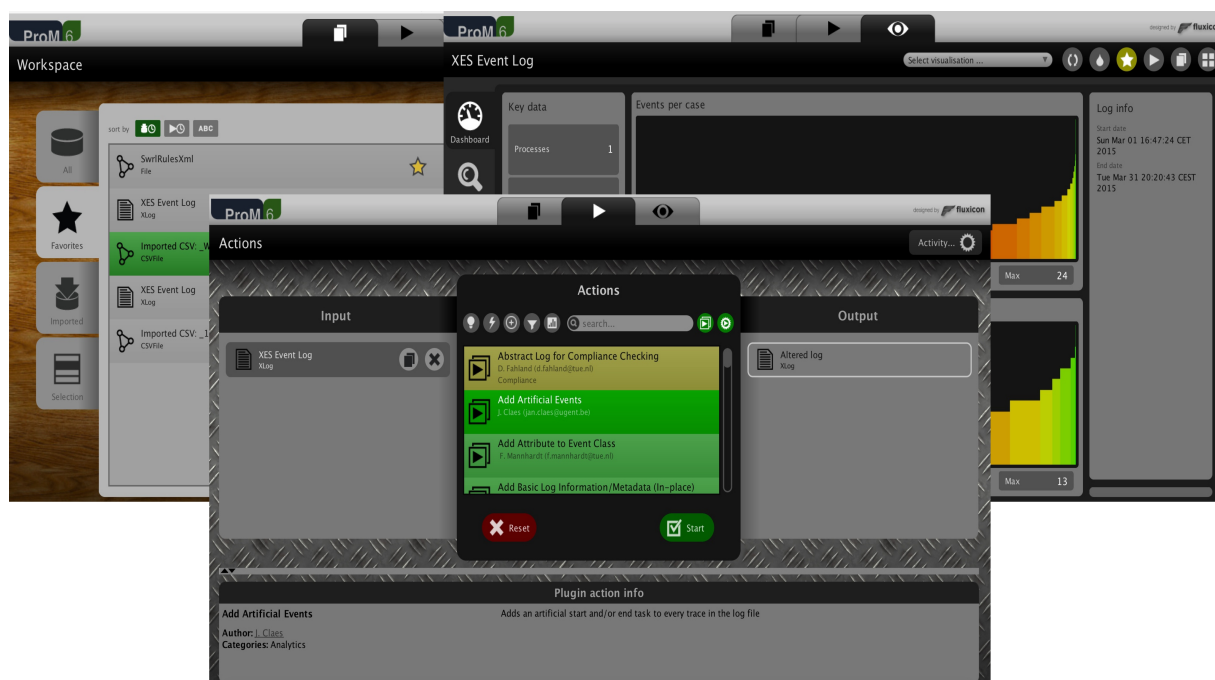
Navrhovaná metodika spočíva v sledovaní dodržania postupnosti aktivít podľa postupnosti zachytenej v definovanom formálnom zápise pravidiel. Vyžadovaný predchodca (aktivita alebo element)

aktuálne analyzovanej aktivity nemusí byť priamym predchodcom ale k jej výskytu mohlo dôjsť v ľubovoľnom čase postupnosti. Predpokladá sa splnená podmienka výskytu aktivity v rámci práve analyzovanej inštancie skúmaného prípadu. Jednotlivé udalosti zo spracovávaného logu udalostí sú automaticky zoskupené do sekvencií, ktoré odrážajú skutočný pohyb užívateľa v systéme. Tieto postupnosti je možné následne verifikovať postupne voči zadaným pravidlám a je možné zaznamenať, ktoré aktivity v danej inštancii analyzovaného procesu, porušujú konkrétne pravidlo.

11.3.4. Toolchain

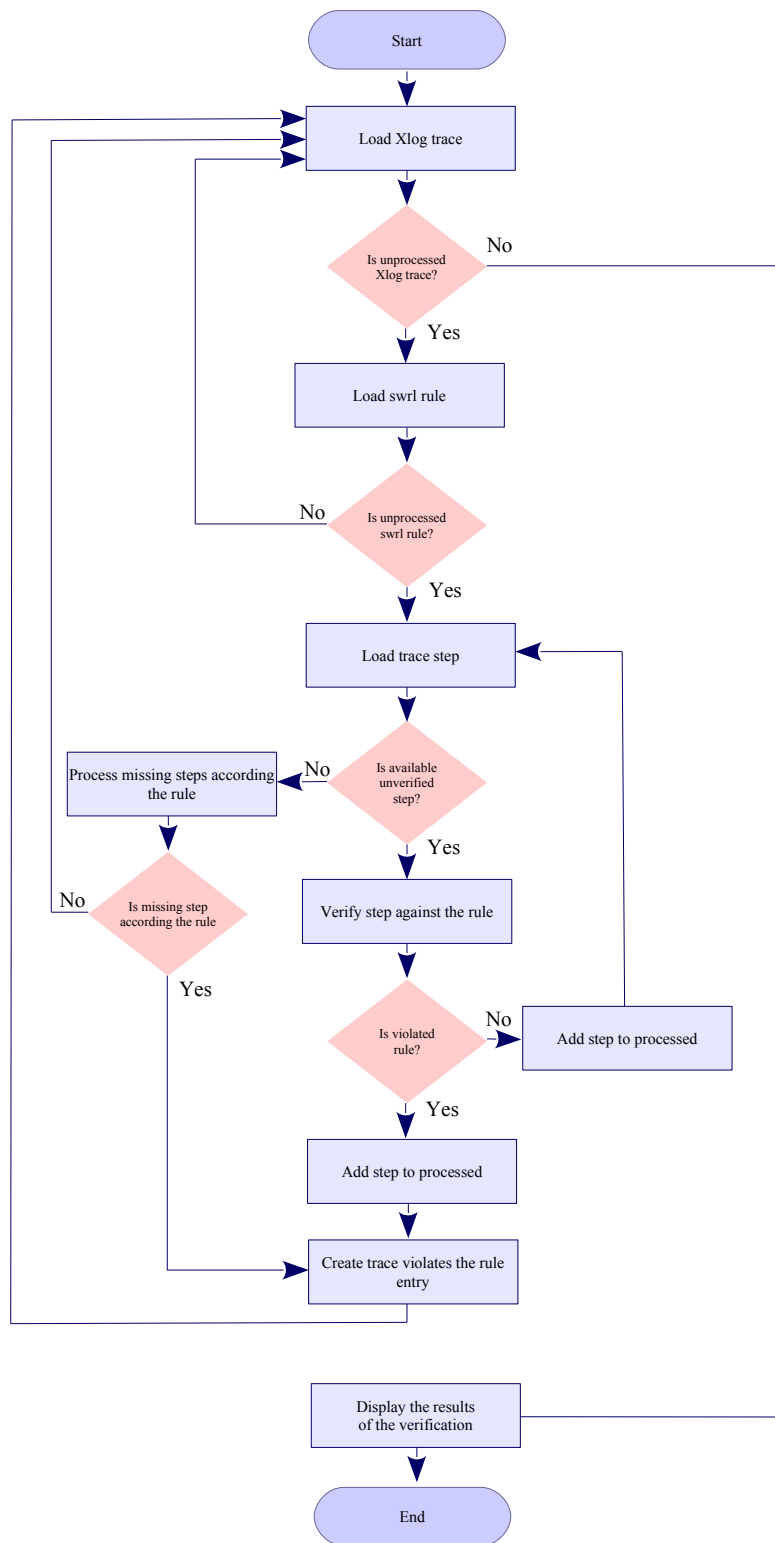
Pri vypracovaní práce bude nutné použiť niekoľko softwarových nástrojov.

- **Softwarové ecommerce riešenie** ako poskytovateľ heatmapy, ktorá slúži ako zdroj dát
- **Protege** – nástroj na tvorbu ontológií vrátane SWRL pravidiel
- **ProM** – softwarový nástroj, ktorý ponúka rozsiahle množstvo implementovaných metódik process miningu vo forme pluginov. Nástroj je použiteľný zdarma a je postavený na technológii Java. Okolo nástroja je vytvorená malá komunita výskumníkov, ktorí postupne rozširujú funkcionality nástroja implementáciou nových pluginov [7].



Obrázok 22. ProM

11.4. Implementácia

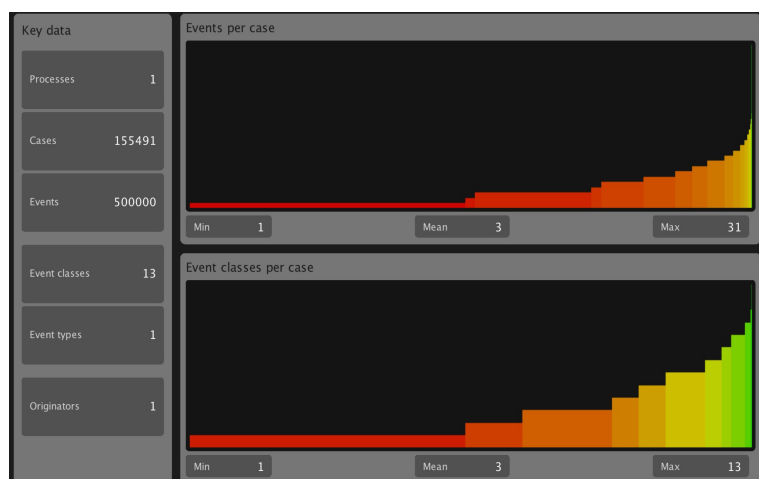


Obrázok 23. Schéma verifikácie prípadov zachytených v logu udalostí. Jednotlivé prípady sú typu Xtrace.

11.5. Výsledky verifikácie s využitím implementovaného pluginu

11.5.1 Konvertovanie heatmapy do formátu XES

- V prvom kroku prekonvertujeme CSV s dátami analyzovanej heatmapy do štandardizovaného formátu, určeného pre zápis logov udalostí, XES.
- CSV obsahujúce vzorek dát z heatmapy obsahuje 500 000 záznamov
- Konverzia dát do štandardizovaného logu udalostí vytvára priestor na doplňujúcu analýzu dát



Obrázok 24. ProM. Analýza logu udalostí - analýza početnosti udalostí v rámci jednotlivých inštancií procesu

- Na Obrázku 24 je možné sledovať početnosť udalostí v rámci jednotlivých inštancií bežiaceho procesu. Prvý graf prezentuje fakt, že maximálna zaznamenaná inštancia procesu je zložená z 31 udalostí. Znamená to, že zákazník po príchode do eshopu pravdepodobne nejakú dobu prechádzal jednotlivými kategóriami produktov a až po nájdení všetkých hľadaných produktov riadne dokončil objednávkový proces. Druhý graf prezentuje počet unikátnych udalostí v rámci jedného prípadu.
- Obrázok 25. a Obrázok 26. nám ukazujú početnosť výskytu jednotlivých udalostí. Obrázok 26. zobrazuje informácie o tom, ktoré udalosti boli v koľkých prípadoch ukončovacou udalosťou daného prípadu. udalosť `a_order_finish`, ktorá v analyzovanom procese vystupuje ako riadna ukončovacia aktivita daného procesu sa v zaznamenaných prípadoch vyskytla len v 3.16% prípadov. To znamená, že len 3.16% návštevníkov eshopu riadne prešlo všetkými krokmi analyzovaného scenára a dokončili objednávkový proces.

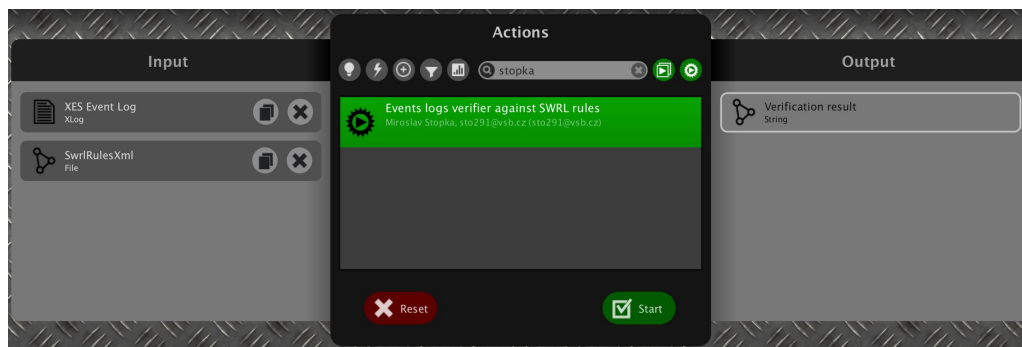
Class	Occurrences (absolute)	Occurrences (relative)
a_incoming	155491	31.098%
a_category	92078	18.416%
a_pl_detail	69790	13.958%
a_search	41954	8.391%
a_pd_buy	39476	7.895%
a_bf_into_basket	30882	6.176%
a_basket_to_order	19723	3.945%
a_article	18555	3.711%
a_order_payment	7495	1.499%
a_order_transport	7495	1.499%
a_filtr	6198	1.24%
a_producer	5946	1.189%
a_order_finish	4917	0.983%

Obrázok 25. ProM. Analýza logu udalostí - analýza výskytu udalostí

End events		
Total number of classes: 12		
Class	Occurrences (absolute)	Occurrences (relative)
a_incoming	76132	48.962%
a_pl_detail	22370	14.387%
a_category	17688	11.376%
a_basket_to_order	11468	7.375%
a_bf_into_basket	7000	4.502%
a_article	6533	4.202%
a_order_finish	4917	3.162%
a_filtr	3154	2.028%
a_producer	2917	1.876%
a_order_payment	2114	1.36%
a_pd_buy	734	0.472%
a_order_transport	464	0.298%

Obrázok 26: ProM. Analýza logu udalostí - analýza koncových udalostí zachytených prípadov

11.5.2 Spustenie implementovaného pluginu

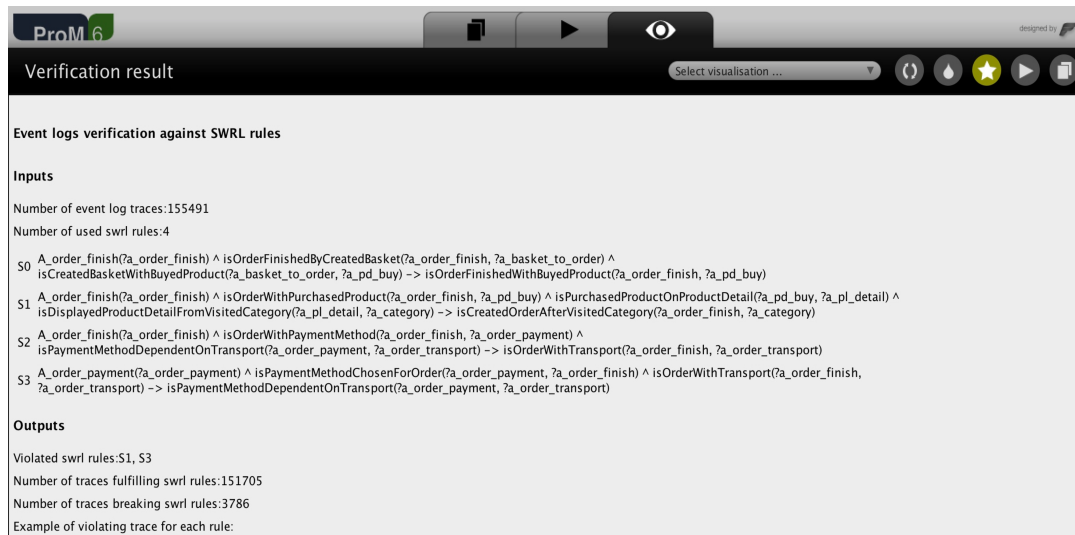


Obrázok 27. ProM. Voľba implementovaného pluginu. Vľavo je zobrazený zoznam požadovaných vstupov. Vpravo je zobrazený zoznam výstupov pluginu.

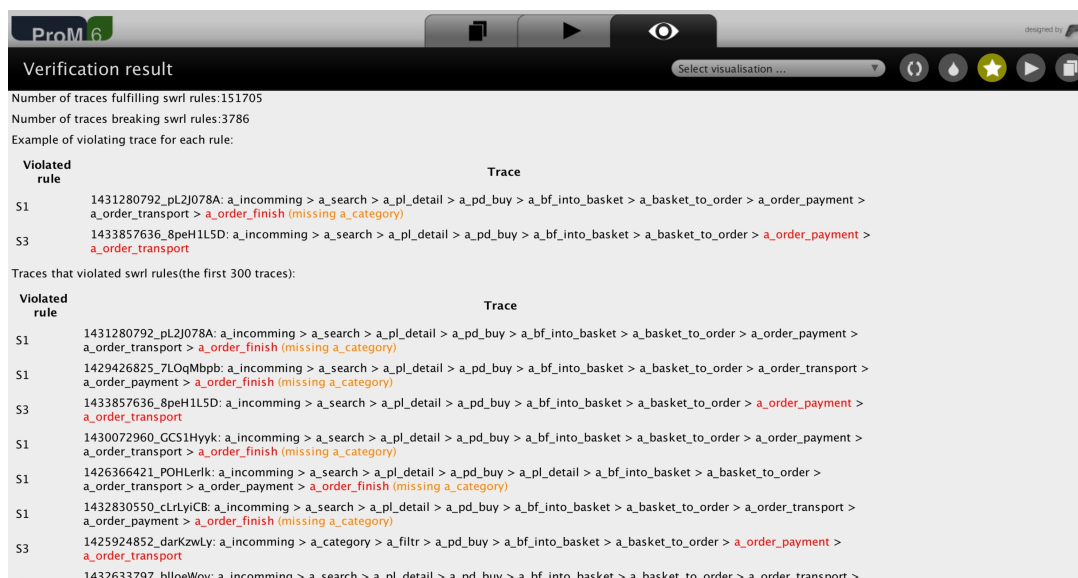
- Po zvolení implementovaného pluginu je nutné zvoliť vstupné dátové zdroje. V našom prípade je

prvým zdrojom vytvorený log udalostí vo formáte XES a druhým zdrojom pravidlá vyexportované z nástroja Protege v súbore formátu XML.

11.5.3 Výsledok verifikácie



Obrázok 28. ProM. Výsledok verifikácie 1



Obrázok 29. ProM. Výsledok verifikácie 2

- Obrázok 28. a Obrázok 29. prezentujú výsledky verifikácie vykonanej s využitím implementovaného pluginu. Na prvom obrázku je možné vidieť počet zanalyzovaných prípadov: 155 491 a počet pravidiel spolu s výpisom týchto pravidiel, voči ktorým boli verifikované jednotlivé prípady. Pravidlá sú vypísané v rovnakom formáte ako sú vytvárané v nástroji Protege.

V časti *Výstupy* na prvom obrázku je možné vidieť aj zoznam porušených pravidiel

- Obrázok 29. prezentuje podrobnejšie informácie spojené s konkrétnymi prípadmi, ktoré porušili nejaké definované pravidlo. Z počtu 155 491 prípadov je počet prípadov, ktoré spĺňajú všetky definované pravidlá 151 705. Počet prípadov, u ktorých bolo zistené porušenie aspoň jedného pravidla je 3 786. Z toho vyplýva, že definované pravidlá spĺňa až cca 97% zaznamenaný prípadov. Toto číslo sa môže zdať ako veľmi vysoké. Vytvára sa ale priestor na ďalšie otázky spojené s ďalšími možnými analýzami. Koľko z prípadov, u ktorých došlo skutočne k riadnemu dokončeniu objednávkového procesu, porušilo niektoré z pravidiel? Koľko z prípadov, ktoré sú vyhodnotené ako porušovatelia pravidiel, nebolo len korektne zachytených v logu udalostí, čo spôsobuje skreslenie celkového počtu prípadov, ktoré sú zaznamenané kompletne a sú vyhodnotené ako prípady s odchýlkou? Obrázok 26. prezentuje počet prípadov, v ktorých návštevník opustil systém hneď po vstupe do systému. Týchto prípadov je 76 132. Ak z celkového počtu analyzovaných prípadov odpočítame tento počet získame údaj 79 359, ktorý predstavuje počet prípadov, kedy užívateľ systému pokračoval v procese aj po prvej aktivite. V tomto prípade by už celkový počet prípadov, ktoré porušili aspoň jedno definované pravidlo nie je 3% ale 5%. Ak by sme v postupnej eliminácii nevhodných prípadov postupovali ďalej a zároveň by došlo aj k úprave algoritmu za účelom zmeny cieľa verifikácie – aktuálny algoritmus verifikuje log udalostí za účelom zodpovedania otázok spojených s pôvodným procesným modelom, získali by sme iný pohľad na výsledky analýzy zachytenej reality. Novým cieľom by mohla byť verifikácia logu udalostí za účelom zodpovedania otázok spojených priamo s prípadmi zachytených v logu. V spojení s výsledkami prvej verifikácie by sme pravdepodobne došli zisteniu, že skutočný počet prípadov, u ktorých došlo k riadnemu dokončeniu objednávkového procesu a došlo v týchto prípadoch k porušeniu aspoň jedného pravidla môže byť aj v počte niekoľkých desiatok percent z celkového počtu prípadov, v ktorých došlo k dokončeniu objednávkového procesu.
- Obrázok 29. prezentuje príklad reálneho prípadu, u ktorého došlo k porušeniu aspoň jedného pravidla. Okrem toho je možné sledovať aj konkrétnu príčinu, podľa ktorej je tento prípad ohodnotený ako prípad s odchýlkou.
- Príklad takého prípadu, ktoré porušilo aspoň jedno definované pravidlo je:

1431280792_pL2J078A: a_incomming > a_search > a_pl_detail > a_pd_buy > a_bf_info_basket > a_basket_to_order > a_order_payment > a_order_transport > a_order_finish (missing a_category)

V tomto prípade došlo k riadnemu dokončeniu objednávkového procesu (definované výskytom udalostí *a_order_finish*), prípad ale neobsahuje zachytenú udalosť návštevy konkrétnej kategórie produktov (definovaná udalosťou *a_category*). Podľa pravidla S1 by každá, riadne dokončená, objednávka mala byť výsledkom návštevy nejakej konkrétnej kategórie produktov. V tomto prípade je ale možné vyčítať, že návštevník po príchode do systému použil okamžité funkcionality vyhľadávania a vo výsledku vyhľadávania úspešne našiel produkt, ktorý okamžite zakúpil. K návšteve konkrétnej kategórie teda nedošlo. Ak by sme predpokladali, že jediným cieľom provozovateľa ecommerce systému je len generovanie objednávok, bolo by možné toto narušenie

pravidla ignorovať. Môžeme si ale položiť otázku, aký veľký objem objednávok v porovnaní s celkovým počtom objednávok je vykonaných práve takýmto spôsobom? V prípade, že bude takto vytvorených objednávok viac ako 50%, môže každý zásah v časti systému vyhľadávania viesť k výraznému zníženiu počtu objednávok už len z dôvodu zmeny grafického rozhrania, ktorej provozovateľ nebude spočiatku prikladať veľkú dôležitosť a pozornosť. A naopak zo optimalizovaním, takto často využívanej, vyhľadávacej sekcie je možné dosiahnuť ďalší nárast návštevníkov systému. Vďaka analýze prípadov objednávkového procesu je možné určiť, ktoré sú zdrojom najvýraznejších príjmov a pomocou návrhov vyplývajúcich z tejto analýzy je možné predchádzať výrazným škodám.

- ďalším príkladom prípadu, ktorý porušil aspoň jedno definované pravidlo je:

1433857636_8peHIL5D: a_incomming > a_search > a_pl_detail > a_pd_buy > a_bf_into_basket > a_basket_to_order > a_order_payment > a_order_transport

V tomto prípade došlo k porušeniu pravidla S3, ktoré hovorí o závislosti platobnej metóde na zvolenej doprave. V ecommerce systémoch je žiadúce, aby nebolo možné zvoliť ľubovoľný typ platobnej metódy bez ohľadu na dopravu. Názorným príkladom môže byť situácia, kedy zvolí zákazník platobnú metódu *Dobierka* ale ako spôsob dopravy zvolí *Osobný odber*. V zaznamenanom prípade *1433857636_8peHIL5D* došlo k situácii, kedy užívateľ zvolil najprv platobnú metódu a až následne spôsob dopravy. Tento stav vytvára priestor na nežiaducu situáciu v budúcnosti. Je nutné vytvoriť návrh na zmenu, ktorá bude reagovať na riziko vzniku tohto stavu.

- Obrázok 29. nakoniec zobrazuje vzorku prípadov, ktoré porušili aspoň jedno definované pravidlo

11.5.3 Záver a návrhy na zmenu

Z pôvodne položených otázok v kapitole sme následnou analýzou vyselectovali otázky, na ktoré je možné získať odpovede verifikáciou pôvodného procesného modelu:

1. Odpovedá dostupný procesný model skutočnému stavu reality? - *pravidlo nie je možné definovať*
2. Je pravda, že objednávku je možné dokončiť až po vložení nejakého produktu do košíku? - *pravidlo S0*
3. Je zvolený druh platobnej metódy závislý na zvolenej doprave? - *pravidlo S3*
4. Je možné vykonať objednávku bez návštevy kategórie produktov? - *pravidlo S1*
5. Je pravda, že každá dokončená objednávka obsahuje dopravu? - *pravidlo S2*

Následnou verifikáciou logov udalostí, v ktorých je zachytený skutočný stav reality prostredníctvom prípadov inštancií analyzovaného procesu, sme zistili porušenie pravidiel S1 a S3.

- Porušenie pravidla S3 znamená, že je možné vysloviť tvrdenie:

Nie je pravda, že zvolený druh platobnej metódy je závislý na zvolenej doprave

Na základe analýzy popísanej v predchádzajúcej kapitole, ktorá sa zachytáva aj možné nástrahy vyplývajúce z porušenia tohto pravidla, je možné vysloviť návrh na úpravu systému, ktorá bude viesť k eliminácii faktorov, ktoré vedú k vytvoreniu priestoru na výskyt tejto odchýlky. Navrhujeme teda zmenu v aplikácii, konkrétne zmenu v implementácii objednávkového formulára, ktorá bude vykonaná takým spôsobom, že nebude možné zvoliť platobnú metódu pred tým ako dôjde k voľbe spôsobu dopravy.

- Pravidlo S1 vzniklo ako negácia otázky číslo 4., ktorá znie *Je možné vykonať objednávku bez návštevy kategórie produktov?* Preklad pravidla je možné uviesť ako *Je pravda, že nie je možné vykonať objednávku bez návštevy kategórie produktov?*. Po porušení pravidla S1 je teda možné tvrdiť, že *Nie je pravda, že nie je možné objednávku vykonať bez návštevy kategórie produktov*. A teda výsledný výrok znie *Objednávku je možné vykonať bez návštevy kategórie produktov*. Tento výrok je zároveň odpoveďou ANO na otázku číslo 4. V tomto prípade teda porušenie pravidla znamená kladnú odpoveď na položenú otázku.
- Na základe definícií jednotlivých pravidiel je ale možné vyvrátiť aj pravidlo S2, ktoré vo svojej definícii platnosti predpokladá kladné potvrdenie pravidla S3. Keďže došlo k porušeniu pravidla S3, a teda k potvrdeniu neplatnosti pravidla S3, je možné vysloviť tvrdenie, že pravidlo S2 je neplatné.
- Odpoveď na otázku číslo 1 *Odpovedá dostupný procesný model skutočnému stavu reality?*, vyplýva z neplatnosti ostatných pravidiel. Každé definované pravidlo vychádza z analyzovaného modelu. Splnenie všetkých pravidiel znamená, že verifikovaná realita zachytená v logu udalostí spĺňa definované pravidlá a teda analyzovaný model odpovedá verifikovanej realite. Nesplnenie pravidla znamená, že verifikovaná realita nespĺňa všetky definované pravidlá a teda analyzovaný model neodpovedá verifikovanej realite

12. Záver

Obsah diplomovej práce je možné rozdeliť na dve časti. Prvá časť poskytuje súhrn poznatkov z oblasti process miningu. Popisuje hlavné oblasti process miningu – objavovanie, porovnávanie, vylepšovanie, a podrobne popisuje možné perspektívy, prostredníctvom ktorých je možné nahliadnuť na analyzované procesné modely. Okrem toho sa práca podrobne venuje jednotlivým krokom metodík process miningu od problematiky zberu dát, cez otázky a problémy spojené so spôsobom vizualizácie alebo spôsoby vyhodnocovania kvality. Prvá časť práce je zakončená aplikácií postupov process miningu v oblasti operačného supportu.

V druhej časti práce je popísané riešenie konkrétneho problému analýzy procesného modelu z oblasti ecommerce. Pri riešení som postupoval podľa všetkých pokynov vyplývajúcich z problematiky process miningu. Boli položené konkrétne otázky vyplývajúce z analyzovaného modelu, vykonal som zber dát logov udalostí, ktoré zachytávajú skutočný stav reality, v spolupráci s vedúcim práce bola navrhnutá nová metodika analýzy procesného modelu, ktorú som realizoval prostredníctvom rozšírenia do nástroja ProM. Pomocou nového rozšírenia je možné ohodnotiť stav pôvodného modelu, ktorý je vyjadrený prostredníctvom konkrétnych otázok, voči realite. Hodnotenie prebieha verifikáciou prípadov zachytených v logu udalostí oproti položeným otázkam

Zhodnotenie úspešnosti a použiteľnosti navrhutej metodiky verifikácie procesného modelu ako aj získanie zmysluplných výsledkov je veľmi závislé na položených otázkach a predovšetkým na rozsahu riešenej domény. Pri aplikovaní metodiky na dáta hrubšej granularity alebo pri riešení otázok modelu, ktorý zachytáva vysokú úroveň abstrakcie, môže byť pokladanie správnych otázok a následne aj získanie zmysluplných výsledkov prostredníctvom aplikácie metodiky veľmi problematické. Z výsledkov experimentov s existujúcimi metodikami vyplýva, že aplikácia existujúcich postupov na dáta a typ domény, pre ktoré nie je použitá metodika priamo vytvorená, môže poskytovať len čiastočné výsledky. Pravdepodobne existujú prípady, v ktorých získanie len čiastočných odpovedí môže v širšej perspektíve znamenať splnenie zadanej úlohy. V prípade potreby analýzy modelu vyššej úrovne abstrakcie by som túto metodiku pravdepodobne nezvolil a analyzovaný model by som rozdelil na modely zachytávajúce menšie úseky celkového procesu, ktoré by som následne v podobe petriho sietí verifikoval spúšťaním jednotlivých prípadov. S využitím tejto techniky je zároveň možné v reálnom čase zobrazovať návrhy na zmenu v modeli na základe spustených prípadov. V našom prípade sme ale vďaka návrhu metodiky priamo pre riešenie problematiky úspešne získali kompletne odpovede, na základe ktorých boli popísané návrhy na zmenu v systéme ako aj návrhy na zmenu v analyzovanom procesnom modeli.

Použitá literatura

- [1] Wil van der Aalst. *Process mining: Data science in action Second edition*. Springer, 2013 ISBN 978-3-662-49851-4.
- [2] J.C.A.M. Buijs, B.F. Van Dongen, W.M.P Van Der Aalst. *Quality dimensions in process discovery: The importance of fitness, precision, generalization and simplicity [online]*. 2013.
- [3] Ophir Frieder, David A. Grossman. *Information Retrieval: Algorithms and Heuristics (2nd edition)*. Springer, 2004, ISBN-10 1402030045.
- [4] Reisig Wolfgang: *Understanding Petri Nets*. Springer-Verlag Berlin Heidelberg, 2013, ISBN 978-3-642-33277-7.
- [5] J.C.A.M. Buijs. *Mapping data sources to XES in a generic way*. 2010, 123s, Master Thesis, Supervisors: W.M.P Van Der Aalst, H.M.W. Verbeek, G.H.L. Fletcher. Technische Universiteit Eindhoven.
- [6] W.M.P Van Der Aalst. *Event Logs, What kind of data does process mining require?[online]*. 2016.
- [7] H.M.W. Eric Verbeek, R.P. Jagadeesh Chandra Bose. *Prom 6 Tutorial[online]*. 2010.
- [8] Andrea Burattin. *Process Mining Techniques in Business Environments: Theoretical Aspects, Algorithms, Techniques and Open Challenges in Process Mining*. Springer, 2015 ISBN 978-3319174815
- [9] R.S. Mans, M.H. Schonenberg, M. Song, W.M.P Van Der Aalst. *Process mining in Healthcare[online]*. 2008, Case Study.

Zoznam príloh

A. ProM with implemented extension

- Elektronická príloha
- Obsahuje aplikáciu ProM doplnenú o rozšírenie implementujúce verifikáciu procesného modelu s využitím navrhutej metodiky
- Aplikácia ProM je doplnená aj o implementovaný parser SWRL pravidiel vygenerovaných v nástroji Protege

B. Verification inputs

- Elektronická príloha
- Obsahuje vstupné dáta pre implementované rozšírenie

C. Video

- Elektronická príloha
- Video, ktoré zachytáva použitie implementovaného rozšírenia